REGULARIZATION OF INVERSE PROBLEMS

LECTURE NOTES

Christian Clason

January 2, 2020

christian.clason@uni-due.de
https://udue.de/clason

CONTENTS

PREFACE

Inverse problems occur wherever a quantity cannot be directly measured but only inferred through comparing observations to the output of mathematical models. Examples of such problems are ubiquitous in biomedical imaging, non-destructive testing, and calibration of financial models. The name *inverse problem* is due to the fact that it contains as a *direct problem* the evaluation of the model given an estimate of the sought-for quantity. However, it is more relevant from a mathematical point of view that such problems are *ill-posed* and cannot be treated by standard methods for solving (non)linear equations.[1]

The mathematical theory of inverse problems is therefore a part of functional analysis: in the same way that the latter is concerned with the question when an equation $F(x) = y$ between infinite-dimensional vector spaces admits a unique solution x that depends continuously on y, the former is concerned with conditions under which this is *not* the case and with methods to at least obtain a reasonable approximation to x. In finite dimensions, this essentially corresponds to the step from regular to inconsistent, under- or overdetermined, or ill-conditioned systems of linear equations.

Although inverse problems are increasingly studied in Banach spaces, we will restrict ourselves in these notes to Hilbert spaces as here the theory is essentially complete and allows for full characterizations in many cases. We will also not consider statistical (frequentist or Bayesian) inverse problems, which have also become prominent in recent years. For the former, we refer to [Scherzer et al. 2009; Ito & Jin 2014; Schuster et al. 2012]; the latter is still missing a broad and elementary exposition aimed at a mathematical audience.

These notes are based on graduate lectures given from 2014–2019 at the University of Duisburg-Essen. As such, no claim is made of originality (beyond possibly the selection and arrangement of the material). Rather, like a magpie, I have tried to collect the shiniest results and proofs I could find. Here I mainly followed the seminal work [Engl, Hanke & Neubauer 1996] (with simplifications by considering only compact instead of bounded linear operators in Hilbert spaces), with additional material from [Hohage 2002; Kindermann 2011; Andreev et al. 2015; Kirsch 2011; Ito & Jin 2014; Kaltenbacher, Neubauer & Scherzer 2008]. Further literature consulted during the writing of these notes was [Louis 1989; Hofmann 1999; Rieder 2003; von Harrach 2014; Burger 2007].

[1] Otherwise there would not be need of a dedicated lecture. In fact, a more fitting title would have been *Ill-posed Problems*, but the term *inverse problem* has become widely accepted, especially in applications.

Part I

BASIC FUNCTIONAL ANALYSIS

1 LINEAR OPERATORS IN NORMED SPACES

In this and the following chapter, we collect the basic concepts and results (and, more importantly, fix notations) from linear functional analysis that will be used throughout these notes. For details and proofs, the reader is referred to the standard literature, e.g., [Alt 2016; Brezis 2010].

1.1 NORMED VECTOR SPACES

In the following, X will denote a vector space over the field $\mathbb{K}$, where we restrict ourselves to the case $\mathbb{K} = \mathbb{R}$. A mapping $\|\cdot\| : X \to \mathbb{R}^+ := [0, \infty)$ is called a *norm* (on X) if for all $x \in X$ there holds

(i) $\|\lambda x\| = |\lambda|\|x\|$ for all $\lambda \in \mathbb{K}$,

(ii) $\|x + y\| \leq \|x\| + \|y\|$ for all $y \in X$,

(iii) $\|x\| = 0$ if and only if $x = 0 \in X$.

Example 1.1. (i) Norms on $X = \mathbb{R}^N$ are defined by

$$\|x\|_p = \left(\sum_{i=1}^N |x_i|^p\right)^{1/p}, \qquad 1 \leq p < \infty,$$
$$\|x\|_\infty = \max_{i=1,\dots,N} |x_i|.$$

(ii) Norms on $X = \ell^p$ (the space of real-valued sequences for which the corresponding terms are finite) are defined by

$$\|x\|_p = \left(\sum_{i=1}^\infty |x_i|^p\right)^{1/p}, \qquad 1 \leq p < \infty,$$
$$\|x\|_\infty = \sup_{i=1,\dots,\infty} |x_i|.$$

(iii) Norms on $X = L^p(\Omega)$ (the space of real-valued measurable functions on the domain $\Omega \subset \mathbb{R}^d$ for which the corresponding terms are finite) are defined by

$$\|u\|_p = \left(\int_\Omega |u(x)|^p\right)^{1/p}, \qquad 1 \le p < \infty,$$
$$\|u\|_\infty = \operatorname*{ess\,sup}_{x\in\Omega} |u(x)|.$$

(iv) A norm on $X = C(\overline{\Omega})$ (the space of continuous functions on $\overline{\Omega}$) is defined by

$$\|u\|_C = \sup_{x\in\Omega} |u(x)|.$$

Similarly, a norm on the space $C^k(\overline{\Omega})$ of k times continuously differentiable functions is defined by $\|u\|_{C^k} = \sum_{j=0}^k \|u^{(j)}\|_C$.

If $\|\cdot\|$ is a norm on X, the pair $(X, \|\cdot\|)$ is called a *normed vector space*, and one frequently denotes this by writing $\|\cdot\|_X$. If the norm is canonical (as in Example 1.1 (ii)–(iv)), it is often omitted, and one speaks simply of "the normed vector space X".

Two norms $\|\cdot\|_1$, $\|\cdot\|_2$ are called *equivalent* if there exist constants $c_1, c_2 > 0$ such that

$$c_1\|x\|_2 \le \|x\|_1 \le c_2\|x\|_2 \qquad \text{for all } x \in X.$$

If X is finite-dimensional, all norms on X are equivalent. However, in this case the constants c_1, c_2 may depend on the dimension of X; in particular, it may be the case that $c_1(N) \to 0$ or $c_2(N) \to \infty$ for $\dim X = N \to \infty$, making the corresponding inequality useless for growing dimensions. Avoiding such dimension-dependent constants is therefore one of the main reasons for studying inverse problems in infinite-dimensional spaces.

If $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are normed vector spaces with $X \subset Y$, then X is called *continuously embedded* in Y, denoted by $X \hookrightarrow Y$, if there exists a $C > 0$ such that

$$\|x\|_Y \le C\|x\|_X \qquad \text{for all } x \in X.$$

A norm directly induces a notion of convergence, the so-called *strong convergence*: A sequence $\{x_n\}_{n\in\mathbb{N}} \subset X$ converges (strongly in X) to a $x \in X$, denoted by $x_n \to x$, if

$$\lim_{n\to\infty} \|x_n - x\|_X = 0.$$

A set $U \subset X$ is called

- *closed* if for every convergent sequence $\{x_n\}_{n\in\mathbb{N}} \subset U$ the limit $x \in X$ lies in U as well;

- *compact* if every sequence $\{x_n\}_{n\in\mathbb{N}} \subset U$ contains a convergent subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ with limit $x \in U$;
- *dense* in X if for all $x \in X$ there exists a sequence $\{x_n\}_{n\in\mathbb{N}} \subset U$ with $x_n \to x$.

The union of U with the set of all limits of convergent sequences in U is called the *closure* $\overline{U}$ of U; obviously, U is dense in $\overline{U}$.

A normed vector space X is called *complete*, if every Cauchy sequence in X converges; in this case, X is called a *Banach space*. All spaces in Example 1.1 are Banach spaces. If X is an incomplete normed space, we denote by $\overline{X}$ the *completion* of X (with respect to the norm $\|\cdot\|_X$).

Finally, we define for later use for given $x \in X$ and $r > 0$

- the *open ball* $U_r(x) := \{z \in X \mid \|x - z\|_X < r\}$ and
- the *closed ball* $B_r(x) := \{z \in X \mid \|x - z\|_X \leq r\}$.

The closed ball around $x = 0$ with radius $r = 1$ is also referred to as the *unit ball* B_X. A set $U \subset X$ is called

- *open* if for all $x \in U$ there exists an $r > 0$ such that $U_r(x) \subset U$ (i.e., all $x \in U$ are *interior points* of U);
- *bounded* if it is contained in a closed ball $B_r(0)$ for an $r > 0$;
- *convex* if for all $x, y \in U$ and $\lambda \in [0,1]$ also $\lambda x + (1-\lambda)y \in U$.

In normed spaces, the complement of an open set is also closed and vice versa (i.e., the closed sets in the sense of topology are exactly the (sequentially) closed sets in the sense of the above definition). The definition of a norm directly implies that open and closed balls are convex. On the other hand, the unit ball is compact if *and only if* X is finite-dimensional; this will be of fundamental importance throughout these notes.

1.2 BOUNDED OPERATORS

We now consider mappings between normed vector spaces. In the following, let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed vector spaces, $U \subset X$, and $F : U \to Y$ be a mapping. We denote by

- $\mathcal{D}(F) := U$ the *domain* of F;
- $\mathcal{N}(F) := \{x \in U \mid F(x) = 0\}$ the "kernel" or "null space" of F;
- $\mathcal{R}(F) := \{F(x) \in Y \mid x \in U\}$ the "range" of F.

We call $F : U \to Y$

- *continuous* in $x \in U$ if for all $\varepsilon > 0$ there exists a $\delta > 0$ with
$$\|F(x) - F(z)\|_Y \leq \varepsilon \qquad \text{for all } z \in U \text{ with } \|x - z\|_X \leq \delta;$$
- *Lipschitz continuous* if there exists a *Lipschitz constant* $L > 0$ with
$$\|F(x_1) - F(x_2)\|_Y \leq L\|x_1 - x_2\|_X \qquad \text{for all } x_1, x_2 \in U.$$

A mapping $F : X \to Y$ is thus continuous if and only if $x_n \to x$ implies $F(x_n) \to F(x)$; it is *closed* if both $x_n \to x$ and $F(x_n) \to y$ imply $F(x) = y$.

If $F : X \to Y$ is linear (i.e., $F(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 F(x_1) + \lambda_2 F(x_2)$ for all $\lambda_1, \lambda_2 \in \mathbb{R}$ and $x_1, x_2 \in X$), continuity of F is equivalent to the existence of a $C > 0$ such that
$$\|F(x)\|_Y \leq C\|x\|_X \qquad \text{for all } x \in X.$$

For this reason, continuous linear mappings are called *bounded*; one also speaks of a bounded linear *operator*. (In the following, we generically denote these by T and omit the parentheses around the argument to indicate this.) If Y is complete, the vector space $\mathbb{L}(X, Y)$ of bounded linear operators becomes a Banach space when endowed with the *operator norm*
$$\|T\|_{\mathbb{L}(X,Y)} = \sup_{x \in X \setminus \{0\}} \frac{\|Tx\|_Y}{\|x\|_X} = \sup_{\|x\|_X \leq 1} \|Tx\|_Y = \sup_{\|x\|_X = 1} \|Tx\|_Y,$$
which is equal to the minimal constant C in the definition of continuity. This immediately implies that
$$\|Tx\|_Y \leq \|T\|_{\mathbb{L}(X,Y)} \|x\|_X \qquad \text{for all } x \in X.$$

As in linear algebra, we call T

- *injective* if $\mathcal{N}(T) = \{0\}$;
- *surjective* if $\mathcal{R}(T) = Y$;
- *bijective* if T is injective and surjective.

If $T \in \mathbb{L}(X, Y)$ is bijective, the *inverse* $T^{-1} : Y \to X$, $Tx \mapsto x$, is continuous if and only if there exists a $c > 0$ with

$$c\|x\|_X \leq \|Tx\|_Y \qquad \text{for all } x \in X; \tag{1.1}$$

in this case, $\|T^{-1}\|_{\mathbb{L}(Y,X)} = c^{-1}$ holds for the maximal c satisfying (1.1). The question of when this is the case is answered by the following three main theorems of functional analysis (that all are more or less direct consequences of the Open Mapping Theorem).

Theorem 1.2 (continuous inverse). *If X, Y are Banach spaces and $T \in \mathbb{L}(X, Y)$ is bijective, then $T^{-1} : Y \to X$ is continuous.*

Of particular relevance for inverse problems is the situation that T is injective but not surjective; in this case, one would like to at least have a continuous inverse on the range of T. However, this does not hold in general, which is one of the fundamental issues in infinite-dimensional inverse problems.

Theorem 1.3 (closed range). *If X, Y are Banach spaces and $T \in \mathbb{L}(X, Y)$ is injective, then $T^{-1} : \mathcal{R}(T) \to X$ is continuous if and only if $\mathcal{R}(T)$ is closed.*

The following theorem completes the trio.

Theorem 1.4 (closed graph). *Let X, Y be Banach spaces. Then $T : X \to Y$ is continuous if and only if T is closed.*

We now consider sequences of linear operators. Here we distinguish two notions of convergence: A sequence $\{T_n\}_{n\in\mathbb{N}} \subset \mathbb{L}(X, Y)$ converges to $T \in \mathbb{L}(X, Y)$

(i) *pointwise* if $T_n x \to Tx$ (strongly in Y) for all $x \in X$;

(ii) *uniformly* if $T_n \to T$ (strongly in $\mathbb{L}(X, Y)$).

Obviously, uniform convergence implies pointwise convergence; weaker conditions are provided by another main theorem of functional analysis.

Theorem 1.5 (Banach–Steinhaus). *Let X be a Banach space, Y be a normed vector space, and $\{T_i\}_{i\in I} \subset \mathbb{L}(X, Y)$ be a family of pointwise bounded operators, i.e., for all $x \in X$ there exists an $M_x > 0$ with $\sup_{i\in I} \|T_i x\|_Y \le M_x$. Then*

$$\sup_{i\in I} \|T_i\|_{\mathbb{L}(X,Y)} < \infty.$$

Corollary 1.6. *Let X, Y be Banach spaces and $\{T_n\}_{n\in\mathbb{N}} \subset \mathbb{L}(X, Y)$. Then the following statements are equivalent:*

(i) $\{T_n\}_{n\in\mathbb{N}}$ converges uniformly on compact subsets of X;

(ii) $\{T_n\}_{n\in\mathbb{N}}$ converges pointwise on X;

(iii) $\{T_n\}_{n\in\mathbb{N}}$ converges pointwise on a dense subset $U \subset X$ and

$$\sup_{n\in\mathbb{N}} \|T_n\|_{\mathbb{L}(X,Y)} < \infty.$$

Corollary 1.7. *Let X, Y be Banach spaces and $\{T_n\}_{n\in\mathbb{N}} \subset \mathbb{L}(X, Y)$. If T_n converges pointwise to a $T : X \to Y$, then T is bounded.*

2 COMPACT OPERATORS IN HILBERT SPACES

As mentioned in the preface, the theory of linear inverse problems can be stated most fully in Hilbert spaces. There, the analogy to ill-conditioned linear systems of equations is also particularly evident.

2.1 INNER PRODUCTS AND WEAK CONVERGENCE

Hilbert spaces are characterized by an additional structure: a mapping $(\cdot\,|\,\cdot) : X \times X \to \mathbb{R}$ on a normed vector space X over the field $\mathbb{R}$ is called *inner product* if

(i) $(\alpha x + \beta y \,|\, z) = \alpha\,(x \,|\, z) + \beta\,(y \,|\, z)$ for all $x, y, z \in X$ and $\alpha, \beta \in \mathbb{R}$;

(ii) $(x \,|\, y) = (y \,|\, x)$ for all $x, y \in X$;

(iii) $(x \,|\, x) \geq 0$ for all $x \in X$ with $(x \,|\, x) = 0$ if and only if $x = 0$.

An inner product induces a norm

$$\|x\|_X := \sqrt{(x \,|\, x)_X}$$

which satisfies the *Cauchy–Schwarz inequality*

$$|\,(x \,|\, y)_X\,| \leq \|x\|_X \|y\|_X.$$

(If one argument is fixed, the inner product is hence continuous in the other with respect to the induced norm.) If X is complete with respect to the induced norm (i.e., $(X, \|\cdot\|_X)$ is a Banach space), then X is called a *Hilbert space*; if the inner product and hence the induced norm is canonical, it is frequently omitted.

Example 2.1. Example 1.1 (i–iii) for $p = 2$ are Hilbert spaces, where the inner product is defined by

(i) for $X = \mathbb{R}^N$: $\quad (x \,|\, y)_X = \sum_{i=1}^{N} x_i y_i,$

(ii) for $X = \ell^2$: $\quad (x \,|\, y)_X = \sum_{i=1}^{\infty} x_i y_i,$

(iii) for $X = L^2(\Omega)$: $\quad (u \,|\, v)_X = \displaystyle\int_\Omega u(x)v(x)\,dx.$

In all cases, the inner product induces the canonical norm.

The inner product induces an additional notion of convergence: the *weak convergence*. A sequence $\{x_n\}_{n\in\mathbb{N}} \subset X$ converges weakly (in X) to $x \in X$, denoted by $x_n \rightharpoonup x$, if

$$(x_n \,|\, z)_X \to (x \,|\, z)_X \qquad \text{for all } z \in X.$$

This notion generalizes the componentwise convergence in $\mathbb{R}^N$ (choose $z = e_i$, the ith unit vector); hence weak and strong convergence coincide in finite dimensions. In infinite-dimensional spaces, strong convergence implies weak convergence but not vice versa. However, if a sequence $\{x_n\}_{n\in\mathbb{N}}$ converges weakly to $x \in X$ and additionally $\|x_n\|_X \to \|x\|_X$, then x_n converges even strongly to x. Furthermore, the norm is *weakly lower semicontinuous*: If $x_n \rightharpoonup x$, then

$$\|x\|_X \leq \liminf_{n\to\infty} \|x_n\|_X. \tag{2.1}$$

This notion of convergence is useful in particular because the Bolzano–Weierstraß Theorem holds for it (in contrast to the strong convergence) even in infinite dimensions: Every bounded sequence in a Hilbert space contains a weakly convergent subsequence. Conversely, every weakly convergent sequence is bounded.

We now consider linear operators $T \in \mathbb{L}(X, Y)$ between Hilbert spaces X and Y. Of particular importance is the special case $Y = \mathbb{R}$, i.e., the space $\mathbb{L}(X, \mathbb{R})$ of *bounded linear functionals* on X. These can be identified with elements of X.

Theorem 2.2 (Fréchet–Riesz). *Let X be a Hilbert space and $\lambda \in \mathbb{L}(X, \mathbb{R})$. Then there exist a unique $z_\lambda \in X$ with $\|\lambda\|_{\mathbb{L}(X,\mathbb{R})} = \|z_\lambda\|_X$ and*

$$\lambda(x) = (z_\lambda \,|\, x)_X \qquad \textit{for all } x \in X.$$

This theorem allows to define for any linear operator $T \in \mathbb{L}(X, Y)$ an *adjoint operator* $T^* \in \mathbb{L}(Y, X)$ via

$$(T^* y \,|\, x)_X = (Tx \,|\, y)_Y \qquad \text{for all } x \in X, y \in Y,$$

which satisfies

(i) $(T^*)^* = T$;

(ii) $\|T^*\|_{\mathbb{L}(Y,X)} = \|T\|_{\mathbb{L}(X,Y)}$;

(iii) $\|T^*T\|_{\mathbb{L}(X,X)} = \|T\|^2_{\mathbb{L}(X,Y)}$.

If $T^* = T$, then T is called *selfadjoint*.

2.2 ORTHOGONALITY AND ORTHONORMAL SYSTEMS

An inner product induces the notion of *orthogonality*: If X is a Hilbert space, then $x, y \in X$ are called *orthogonal* if $(x \mid y)_X = 0$. For a set $U \subset X$,

$$U^\perp := \{x \in X \mid (x \mid u)_X = 0 \text{ for all } u \in U\}$$

is called the *orthogonal complement* of U in X; the definition immediately implies that $U^\perp$ is a closed subspace. In particular, $X^\perp = \{0\}$. Furthermore, $U \subset (U^\perp)^\perp$. If U is a closed subspace, it even holds that $U = (U^\perp)^\perp$ (and hence $\{0\}^\perp = X$). In this case, we have the *orthogonal decomposition*

$$X = U \oplus U^\perp,$$

i.e., every element $x \in X$ can be represented uniquely as

$$x = u + u_\perp \qquad \text{with} \qquad u \in U,\ u_\perp \in U^\perp.$$

The mapping $x \mapsto u$ defines a linear operator $P_U \in \mathbb{L}(X, X)$, called the *orthogonal projection* on U, which has the following properties:

(i) P_U is selfadjoint;

(ii) $\|P_U\|_{\mathbb{L}(X,U)} = 1$;

(iii) $\operatorname{Id} - P_U = P_{U^\perp}$;

(iv) $\|x - P_U x\|_X = \min_{u \in U} \|x - u\|_X$;

(v) $z = P_U x$ if and only if $z \in U$ and $z - x \in U^\perp$.

If U is not a closed subset, only $(U^\perp)^\perp = \overline{U} \supset U$ holds. Hence, for any $T \in \mathbb{L}(X, Y)$ we have

(i) $\mathcal{R}(T)^\perp = \mathcal{N}(T^*)$ and hence $\mathcal{N}(T^*)^\perp = \overline{\mathcal{R}(T)}$;

(ii) $\mathcal{R}(T^*)^\perp = \mathcal{N}(T)$ and hence $\mathcal{N}(T)^\perp = \overline{\mathcal{R}(T^*)}$.

In particular, the null space of a bounded linear operator is always closed; furthermore, T is injective if and only if $\mathcal{R}(T^*)$ is dense in X.

A set $U \subset X$ whose elements are pairwise orthogonal is called an *orthogonal system*. If in addition

$$(x \mid y)_X = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{else,} \end{cases}$$

for all $x, y \in U$, then U is called an *orthonormal system*; an orthonormal system is called *complete*, if there is no orthonormal system $V \subset X$ with $U \subsetneq V$. Every orthonormal system $U \subset X$ satisfies the *Bessel inequality*

$$\sum_{u\in U} |(x \mid u)_X|^2 \leq \|x\|_X^2 \qquad \text{for all } x \in X, \tag{2.2}$$

where at most countably many terms are not equal to zero. If equality holds in (2.2), then U is called an *orthonormal basis*; in this case, U is complete and

$$x = \sum_{u\in U} (x \mid u)_X\, u \qquad \text{for all } x \in X. \tag{2.3}$$

Every Hilbert space contains an orthonormal basis. If one of them is at most countable, the Hilbert space is called *separable*. The Bessel inequality then implies that the sequence $\{u_n\}_{n\in\mathbb{N}} = U$ converges weakly to 0 (but not strongly due to $\|u_n\|_X = 1$!)

Example 2.3. For $X = L^2((0,1))$, an orthonormal basis is given by $\{u_n\}_{n\in\mathbb{Z}}$ for

$$u_n(x) = \begin{cases} \sqrt{2}\sin(2\pi\, n\, x) & n > 0,\\ \sqrt{2}\cos(2\pi\, n\, x) & n < 0,\\ 1 & n = 0.\end{cases}$$

Finally, every closed subspace $U \subset X$ contains an orthonormal basis $\{u_n\}_{n\in\mathbb{N}}$ for which the orthogonal projection on U can be written as

$$P_U x = \sum_{j=1}^{\infty} \left(x \mid u_j\right)_X u_j.$$

2.3 THE SPECTRAL THEOREM FOR COMPACT OPERATORS

Just as Hilbert spaces can be considered as generalizations of finite-dimensional vector spaces, compact operators furnish an analog to matrices. Here a linear operator $T : X \to Y$ is called *compact* if the image of every bounded sequence $\{x_n\}_{n\in\mathbb{N}} \subset X$ contains a convergent subsequence $\{Tx_{n_k}\}_{k\in\mathbb{N}} \subset Y$. A linear operator T is compact if and only if T maps weakly convergent sequences in X to strongly convergent sequences in Y. (This property is also called *complete continuity*.) We will generically denote compact operators by K.

Obviously, every linear operator with finite-dimensional range is compact. In particular, the *identity* $\mathrm{Id} : X \to X$ – like the unit ball B_X – is compact if *and only if* X is finite-dimensional. Furthermore, the space $\mathcal{K}(X, Y)$ of linear compact operators from X to Y is a closed subspace of $\mathbb{L}(X, Y)$ (and hence a Banach space when endowed with the operator norm). This implies that the limit of any sequence of linear operators with finite-dimensional range is compact. If $T \in \mathbb{L}(X, Y)$ and $S \in \mathbb{L}(Y, Z)$ and at least one of the two is compact, then $S \circ T$ is compact as well. Furthermore, T^* is compact if and only if T is compact (which is known as the Schauder Theorem).

Example 2.4. Canonical examples of compact operators are *integral operators*. We consider for $X = Y = L^2(\Omega)$ with $\Omega = (0, 1)$ and for a given *kernel* $k \in L^2(\Omega \times \Omega)$ the operator $K : L^2(\Omega) \to L^2(\Omega)$ defined pointwise via

$$[Kx](t) = \int_0^1 k(s, t)x(s)\, ds \qquad \text{for almost every } t \in \Omega$$

(where $Kx \in L^2(\Omega)$ by Fubini's Theorem). The Cauchy–Schwarz inequality and Fubini's Theorem immediately yield

$$\|K\|_{\mathbb{L}(X,X)} \leq \|k\|_{L^2(\Omega^2)},$$

which also imply that K is a bounded operator from $L^2(\Omega)$ to $L^2(\Omega)$.

Since $k \in L^2(\Omega^2)$ is in particular measurable, there is a sequence $\{k_n\}_{n\in\mathbb{N}}$ of simple functions (i.e., attaining only finitely many different values) with $k_n \to k$ in $L^2(\Omega^2)$. These can be written as

$$k_n(s, t) = \sum_{i,j=1}^{n} \alpha_{ij} \mathbb{1}_{E_i}(s) \mathbb{1}_{E_j}(t),$$

where $\mathbb{1}_E$ is the characteristic function of the measurable interval $E \subset \Omega$ and E_i are a finite disjoint decomposition of Ω. The corresponding integral operators K_n with kernel k_n by linearity of the integral therefore satisfy

$$\|K_n - K\|_{\mathbb{L}(X,X)} \leq \|k_n - k\|_{L^2(\Omega^2)} \to 0,$$

i.e., $K_n \to K$. Furthermore,

$$[K_n x](t) = \int_0^1 k_n(s, t)x(s)\, ds = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \alpha_{ij} \int_{E_i} x(s)\, ds \right) \mathbb{1}_{E_j}(t)$$

and hence $K_n x$ is a linear combination of the $\{\mathbb{1}_{E_j}\}_{1\leq j\leq n}$. This implies that K is the limit of the sequence $\{K_n\}_{n\in\mathbb{N}}$ of operators with finite-dimensional range and therefore compact.

For the adjoint operator $K^* \in \mathbb{L}(X, X)$, one can use the definition of the inner product on $L^2(\Omega)$ together with Fubini's Theorem to show that

$$[K^* y](s) = \int_0^1 k(s,t)y(t)\,dt \qquad \text{for almost every } s \in \Omega.$$

Hence an integral operator is selfadjoint if and only if the kernel is symmetric, i.e., $k(s,t) = k(t,s)$ for almost every $s, t \in \Omega$.

For example, solution operators to (partial) differential equations or convolution operators – and thus a large class of practically relevant operators – can be represented as integral operators and thus shown to be compact.

The central analogy between compact operators and matrices consists in the fact that compact linear operators have at most countably many eigenvalues (which is not necessarily the case for bounded linear operators). Correspondingly, we have the following variant for the Schur factorization, which will be the crucial tool allowing the thorough investigation of linear inverse problems in Hilbert spaces.

Theorem 2.5 (spectral theorem). *Let X be a Hilbert space and $K \in \mathcal{K}(X, X)$ be selfadjoint. Then there exists a (possibly finite) orthonormal system $\{u_n\}_{n\in\mathbb{N}} \subset X$ and a (in this case also finite) null sequence $\{\lambda_n\}_{n\in\mathbb{N}} \subset \mathbb{R} \setminus \{0\}$ with*

$$Kx = \sum_{n\in\mathbb{N}} \lambda_n \,(x \mid u_n)_X\, u_n \qquad \textit{for all } x \in X.$$

Furthermore, $\{u_n\}_{n\in\mathbb{N}}$ forms an orthonormal basis of $\overline{\mathcal{R}(K)}$.

Setting $x = u_n$ immediately implies that u_n is an eigenvector for the eigenvalue λ_n, i.e., $Ku_n = \lambda_n u_n$. By convention, the eigenvalues are sorted by decreasing magnitude, i.e.,

$$|\lambda_1| \geq |\lambda_2| \geq \cdots > 0.$$

With this ordering, the eigenvalues can also be characterized by the *Courant–Fischer min–max principle*

$$\begin{aligned}
\lambda_n &= \min_{V\subset X} \max_{x\in V} \left\{ (Kx \mid x)_X \;\middle|\; \|x\|_X = 1,\ \dim V^\perp = n-1 \right\} \\
&= \max_{V\subset X} \min_{x\in V} \left\{ (Kx \mid x)_X \;\middle|\; \|x\|_X = 1,\ \dim V = n \right\}.
\end{aligned} \tag{2.4}$$

In particular, $\|K\|_{L(X,X)} = |\lambda_1|$.

Part II

LINEAR INVERSE PROBLEMS

3 ILL-POSED OPERATOR EQUATIONS

We now start our study of operator equations that cannot be solved by standard methods. We first consider a linear operator T between two normed vector spaces X and Y. Following Jacques Hadamard, we call the equation $Tx = y$ *well-posed*, if for all $y \in Y$

(i) there exists an $x \in X$ with $Tx = y$;

(ii) this solution is unique, i.e., $z \neq x$ implies $Tz \neq y$;

(iii) this solution depends continuously on y, i.e., for all $\{x_n\}_{n\in\mathbb{N}}$ with $Tx_n \to y$ we also have $x_n \to x$.

If one of these conditions is violated, the equation is called *ill-posed*.

In practice, a violation of the first two conditions often occurs due to insufficient knowledge of reality and can be handled by extending the mathematical model giving rise to the equation. It can also be handled by extending the concept of a solution such that a generalized solution exists for arbitrary $y \in Y$; if this is not unique, one can use additional information on the sought-for x to select a specific solution. For finite-dimensional Hilbert spaces, this leads to the well-known *least squares method*; since then all linear operators are continuous, the problem is then solved in principle (even if the details and in particular the efficient numerical implementation may still take significant effort). However, in infinite dimensions this is not the case, as the following example illustrates.

Example 3.1. We want to compute for given $y \in Y := C^1([0,1])$ the derivative $x := y' \in C([0,1])$, where we assume that the function y to be differentiated is only given by measurements subject to additive noise, i.e., we only have at our disposal

$$\tilde{y} = y + \eta.$$

In general, we cannot assume that the measurement error η is continuously differentiable; but for the sake of simplicity, we assume that it is at least continuous. In this case, $\tilde{y} \in C([0,1])$ as well, and we have to consider the mapping $x = y' \mapsto y$ as a (linear) operator $T : C([0,1]) \to C([0,1])$. Obviously, condition (i) is then violated. But the problem is not well-posed even if the error is continuously differentiable by coincidence: Consider a sequence $\{\delta_n\}_{n\in\mathbb{N}}$ with $\delta_n \to 0$, choose $k \in \mathbb{N}$ arbitrary, and set

$$\eta_n(t) := \delta_n \sin\left(\frac{kt}{\delta_n}\right)$$

as well as $\tilde{y}_n := y + \eta_n$. Then, $\eta_n \in C^1([0,1])$ and

$$\|\tilde{y}_n - y\|_C = \|\eta_n\|_C = \delta_n \to 0,$$

but

$$\tilde{x}_n(t) := \tilde{y}_n'(t) = y'(t) + k\cos\left(\frac{kt}{\delta_n}\right),$$

i.e., $x := y'$ satisfies

$$\|x - \tilde{x}_n\|_C = \|\eta_n'\|_C = k \qquad \text{for all } n \in \mathbb{N}.$$

Hence the error in the derivative x can (depending on k) be arbitrarily large, even if the error in y is arbitrarily small.

(In contrast, the problem is of course well-posed for $T : C([0,1]) \to C^1([0,1])$, since then $\|\eta_n\|_{C^1} \to 0$ implies by definition that $\|\bar{x} - x_n\|_C \leq \|\eta_n\|_{C^1} \to 0$. The occurring norms thus decide the well-posedness of the problem; these are however usually given by the problem setting. In our example, taking $C^1([0,1])$ as image space implies that besides y also y' is measured – and that is precisely the quantity we are interested in, so that we are no longer considering an inverse problem.)

Note that the three conditions for well-posedness are not completely independent. For example, if $T \in \mathbb{L}(X,Y)$ satisfies the first two conditions, and X and Y are Banach spaces, then T is bijective and thus has by Theorem 1.2 a continuous inverse, satisfying also the third condition.

3.1 GENERALIZED INVERSES

We now try to handle the first two conditions for linear operators between Hilbert spaces by generalizing the concept of solution in analogy to the least squares method in $\mathbb{R}^N$. Let X and Y be Hilbert spaces (which we always assume from now on) and consider for $T \in \mathbb{L}(X,Y)$ the equation $Tx = y$. If $y \notin \mathcal{R}(T)$, this equation has no solution. In this case it is reasonable to look for an $x \in X$ that minimizes the distance $\|Tx - y\|_Y$. On the other hand, if $\mathcal{N}(T) \neq \{0\}$, then there exist infinitely many solutions; in this case, we chose the one with minimal norm. This leads to the following definition.

Definition 3.2. An element $x^\dagger \in X$ is called

(i) *least squares solution* of $Tx = y$ if

$$\|Tx^\dagger - y\|_Y = \min_{z\in X} \|Tz - y\|_Y;$$

(ii) *minimum norm solution* of $Tx = y$ if

$$\|x^\dagger\|_X = \min \{\|z\|_X \mid z \text{ is least squares solution of } Tx = y\}.$$

If T is bijective, $x = T^{-1}y$ is obviously the only least squares and hence minimum norm solution. A least squares solution need not exist, however, if $\mathcal{R}(T)$ is not closed (since in this case, the minimum in the definition need not be attained). To answer the question for which $y \in Y$ a minimum norm solution exists, we introduce an operator – called *generalized inverse* or *pseudoinverse* – mapping y to the corresponding minimum norm solution. We do this by first restriction the domain and range of T such that the operator is invertible and then extending the inverse of the restricted operator to its maximal domain.

Theorem 3.3. *Let $T \in \mathbb{L}(X, Y)$ and set*

$$\tilde{T} := T|_{\mathcal{N}(T)^\perp} : \mathcal{N}(T)^\perp \to \mathcal{R}(T).$$

Then there exists a unique linear extension $T^\dagger$, called Moore–Penrose inverse, *of $\tilde{T}^{-1}$ with*

$$\mathcal{D}(T^\dagger) = \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp, \tag{3.1}$$

$$\mathcal{N}(T^\dagger) = \mathcal{R}(T)^\perp. \tag{3.2}$$

Proof. Due to the restriction to $\mathcal{N}(T)^\perp$ and $\mathcal{R}(T)$, the operator $\tilde{T}$ is injective and surjective, and hence there exists a (linear) inverse $\tilde{T}^{-1}$. Thus $T^\dagger$ is well-defined and linear on $\mathcal{R}(T)$. For any $y \in \mathcal{D}(T^\dagger)$, we obtain by orthogonal decomposition unique $y_1 \in \mathcal{R}(T)$ and $y_2 \in \mathcal{R}(T)^\perp$ with $y = y_1 + y_2$. Since $\mathcal{N}(T^\dagger) = \mathcal{R}(T)^\perp$,

$$T^\dagger y := T^\dagger y_1 + T^\dagger y_2 = T^\dagger y_1 = \tilde{T}^{-1} y_1 \tag{3.3}$$

defines a unique linear extension. Hence $T^\dagger$ is well-defined on its whole domain $\mathcal{D}(T^\dagger)$. □

If T is bijective, we obviously have $T^\dagger = T^{-1}$. However, it is important to note that $T^\dagger$ need not be a *continuous* extension.

In the following, we will need the following properties of the Moore–Penrose inverse.

Lemma 3.4. *The Moore–Penrose inverse $T^\dagger$ satisfies $\mathcal{R}(T^\dagger) = \mathcal{N}(T)^\perp$ as well as the* Moore–Penrose equations *(i)* $TT^\dagger T = T$,

(ii) $T^\dagger T T^\dagger = T^\dagger$,

(iii) $T^\dagger T = \operatorname{Id} - P_\mathcal{N}$,

(iv) $TT^\dagger = (P_{\overline{\mathcal{R}}})|_{\mathcal{D}(T^\dagger)}$,

where $P_\mathcal{N}$ and $P_{\overline{\mathcal{R}}}$ denote the orthogonal projections on $\mathcal{N}(T)$ and $\overline{\mathcal{R}(T)}$, respectively.

Proof. We first show that $\mathcal{R}(T^\dagger) = \mathcal{N}(T)^\perp$. By the definition of $T^\dagger$ and (3.3), we have for all $y \in \mathcal{D}(T^\dagger)$ that

$$T^\dagger y = \tilde{T}^{-1} P_{\overline{\mathcal{R}}} y = T^\dagger P_{\overline{\mathcal{R}}} y \tag{3.4}$$

since $y \in \mathcal{D}(T^\dagger) = \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$ implies that $P_{\overline{\mathcal{R}}} y \in \mathcal{R}(T)$ (and not only in $\overline{\mathcal{R}(T)}$ – this fundamental property will be used repeatedly in the following). Hence $T^\dagger y \in \mathcal{R}(\tilde{T}^{-1}) = \mathcal{N}(T)^\perp$, i.e., $\mathcal{R}(T^\dagger) \subset \mathcal{N}(T)^\perp$. Conversely, $T^\dagger T x = \tilde{T}^{-1}\tilde{T} x = x$ for all $x \in \mathcal{N}(T)^\perp$, i.e., $x \in \mathcal{R}(T^\dagger)$. This shows that $\mathcal{R}(T^\dagger) = \mathcal{N}(T)^\perp$ as claimed.

Ad (iv): For $y \in \mathcal{D}(T^\dagger)$, we have from (3.4) and $\mathcal{R}(T^\dagger) = \mathcal{N}(T)^\perp$ that

$$TT^\dagger y = T\tilde{T}^{-1} P_{\overline{\mathcal{R}}} y = \tilde{T}\tilde{T}^{-1} P_{\overline{\mathcal{R}}} y = P_{\overline{\mathcal{R}}} y$$

since $\tilde{T}^{-1} P_{\overline{\mathcal{R}}} y \in \mathcal{N}(T)^\perp$ and $T = \tilde{T}$ auf $\mathcal{N}(T)^\perp$.

Ad (iii): The definition of $T^\dagger$ implies that $T^\dagger T x = \tilde{T}^{-1} T x$ for all $x \in X$ and hence that

$$T^\dagger T x = \tilde{T}^{-1} T\left(P_\mathcal{N} x + (\operatorname{Id} - P_\mathcal{N})x\right) = \tilde{T}^{-1} T P_\mathcal{N} x + \tilde{T}^{-1}\tilde{T}(\operatorname{Id} - P_\mathcal{N})x = (\operatorname{Id} - P_\mathcal{N})x.$$

Ad (ii): Using (iv) and (3.4) yields for $y \in \mathcal{D}(T^\dagger)$ that

$$T^\dagger T T^\dagger y = T^\dagger P_{\overline{\mathcal{R}}} y = T^\dagger y.$$

Ad (i): Directly from (iii) follows that

$$TT^\dagger T x = T(\operatorname{Id} - P_\mathcal{N})x = Tx - TP_\mathcal{N} x = Tx \qquad \text{for all } x \in X.$$ □

(In fact, the Moore–Penrose equations are an equivalent characterization of $T^\dagger$.)

We can now show that the Moore-Penrose inverse indeed yields the minimum norm solution; in passing, we also characterize the least squares solutions.

Theorem 3.5. *For any $y \in \mathcal{D}(T^\dagger)$, the equation $Tx = y$ admits*

(i) *least squares solutions, which are exactly the solutions of*

$$Tx = P_{\overline{\mathcal{R}}}y; \tag{3.5}$$

(ii) *a unique minimum norm solution $x^\dagger \in X$, which is given by*

$$x^\dagger = T^\dagger y.$$

The set of all least squares solutions is given by $x^\dagger + \mathcal{N}(T)$.

Proof. First, $P_{\overline{\mathcal{R}}}y \in \mathcal{R}(T)$ for $y \in \mathcal{D}(T^\dagger)$ implies that (3.5) admits at least one solution. The optimality of the orthogonal projection further implies that any such solution $z \in X$ satisfies

$$\|Tz - y\|_Y = \|P_{\overline{\mathcal{R}}}y - y\|_Y = \min_{w \in \overline{\mathcal{R}(T)}} \|w - y\|_Y \leq \|Tx - y\|_Y \qquad \text{for all } x \in X,$$

i.e., all solutions of (3.5) are least squares solutions of $Tx = y$. Conversely, any least squares solution $z \in X$ satisfies

$$\|P_{\overline{\mathcal{R}}}y - y\|_Y \leq \|Tz - y\|_Y = \min_{x \in X} \|Tx - y\|_Y = \min_{w \in \overline{\mathcal{R}(T)}} \|w - y\|_Y \leq \|P_{\overline{\mathcal{R}}}y - y\|_Y$$

since $P_{\overline{\mathcal{R}}}y \in \mathcal{R}(T)$ and hence $Tz = P_{\overline{\mathcal{R}}}y$. This shows (i).

The least squares solutions are this exactly the solutions of $Tx = P_{\overline{\mathcal{R}}}y$, which can be uniquely represented as $x = \bar{x} + x_0$ with $\bar{x} \in \mathcal{N}(T)^\perp$ and $x_0 \in \mathcal{N}(T)$. Since T is injective on $\mathcal{N}(T)^\perp$, the element $\bar{x}$ must be independent of x (otherwise $Tx' = T\bar{x}' \neq T\bar{x} = P_{\overline{\mathcal{R}}}y$ for $x' = \bar{x}' + x_0$ with $\bar{x}' \neq \bar{x}$). It then follows from

$$\|x\|_X^2 = \|\bar{x} + x_0\|_X^2 = \|\bar{x}\|_X^2 + 2\,(\bar{x} \mid x_0)_X + \|x_0\|_X^2 = \|\bar{x}\|_X^2 + \|x_0\|_X^2 \geq \|\bar{x}\|_X^2$$

that $x^\dagger := \bar{x} \in \mathcal{N}(T)^\perp$ is the unique minimum norm solution.

Finally, $x^\dagger \in \mathcal{N}(T)^\perp$ and $Tx^\dagger = P_{\overline{\mathcal{R}}}y$ together with Lemma 3.4 (iii) and (ii) imply that

$$x^\dagger = P_{\mathcal{N}^\perp}x^\dagger = (\mathrm{Id} - P_{\mathcal{N}})x^\dagger = T^\dagger T x^\dagger = T^\dagger P_{\overline{\mathcal{R}}}y = T^\dagger T T^\dagger y = T^\dagger y,$$

which shows (ii). ◻

We can give an alternative characterization that will later be useful.

Corollary 3.6. *Let $y \in \mathcal{D}(T^\dagger)$. Then $x \in X$ is a least squares solution of $Tx = y$ if and only if x satisfies the* normal equation

$$T^*Tx = T^*y. \tag{3.6}$$

Is additionally $x \in \mathcal{N}(T)^\perp$, then $x = x^\dagger$.

Proof. Theorem 3.5 (i) states that $x \in X$ is a least squares solution if and only if $Tx = P_{\overline{\mathcal{R}}}y$, which is equivalent to $Tx \in \overline{\mathcal{R}(T)}$ and $Tx - y \in \overline{\mathcal{R}(T)}^\perp = \mathcal{N}(T^*)$, i.e., $T^*(Tx - y) = 0$.

Similarly, Theorem 3.5 (ii) implies that a least squares solution x has minimal norm if and only if $x = T^\dagger y \in \mathcal{N}(T)^\perp$. □

The minimum norm solution $x^\dagger$ of $Tx = y$ is therefore also the solution – and hence, in particular, the least squares solution – of (3.6) with minimal norm, i.e.,

$$x^\dagger = (T^*T)^\dagger T^* y. \tag{3.7}$$

We can therefore characterize $x^\dagger$ as the minimum norm solution of (3.6) as well as of $Tx = y$, which can sometimes be advantageous.

Until now, we have considered the pseudo-inverse of its domain without characterizing this further; this we now catch up on. First, by construction $\mathcal{D}(T^\dagger) = \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$. Since orthogonal complements are always closed,

$$\overline{\mathcal{D}(T^\dagger)} = \overline{\mathcal{R}(T)} \oplus \mathcal{R}(T)^\perp = \mathcal{N}(T^*)^\perp \oplus \mathcal{N}(T^*) = Y,$$

i.e., $\mathcal{D}(T^\dagger)$ is dense in Y. If $\mathcal{R}(T)$ is closed, this implies that $\mathcal{D}(T^\dagger) = Y$ (which conversely implies that $\mathcal{R}(T)$ is closed). Furthermore, for $y \in \mathcal{R}(T)^\perp = \mathcal{N}(T^\dagger)$ the minimum norm solution is always $x^\dagger = 0$. The central question is therefore whether a given $y \in \overline{\mathcal{R}(T)}$ is in fact an element of $\mathcal{R}(T)$. If this always holds, $T^\dagger$ is even continuous. Conversely, the existence of a single $y \in \overline{\mathcal{R}(T)} \setminus \mathcal{R}(T)$ already suffices for $T^\dagger$ not to be continuous.

Theorem 3.7. *Let $T \in \mathbb{L}(X, Y)$. Then $T^\dagger : \mathcal{D}(T^\dagger) \to X$ is continuous if and only if $\mathcal{R}(T)$ is closed.*

Proof. We apply the Closed Graph Theorem 1.4, for which we have to show that $T^\dagger$ is closed. Let $\{y_n\}_{n\in\mathbb{N}} \subset \mathcal{D}(T^\dagger)$ be a sequence with $y_n \to y \in Y$ and $T^\dagger y_n \to x \in X$. Lemma 3.4 (iv) then implies that

$$TT^\dagger y_n = P_{\overline{\mathcal{R}}} y_n \to P_{\overline{\mathcal{R}}} y$$

due to the continuity of the orthogonal projection. It follows from this and the continuity of T that

$$P_{\overline{\mathcal{R}}} y = \lim_{n\to\infty} P_{\overline{\mathcal{R}}} y_n = \lim_{n\to\infty} TT^\dagger y_n = Tx, \tag{3.8}$$

i.e., x is a least squares solution. Furthermore, $T^\dagger y_n \in \mathcal{R}(T^\dagger) = \mathcal{N}(T)^\perp$ also implies that

$$T^\dagger y_n \to x \in \mathcal{N}(T)^\perp$$

since $\mathcal{N}(T)^\perp = \overline{\mathcal{R}(T^*)}$ is closed. By Theorem 3.5 (ii), x is thus the minimum norm solution of $Tx = y$, i.e., $x = T^\dagger y$. Hence $T^\dagger$ is closed.

If $\mathcal{R}(T)$ is now closed, we have that $\mathcal{D}(T^\dagger) = Y$ and thus that $T^\dagger : Y \to X$ is continuous by Theorem 1.4. Conversely, if $T^\dagger$ is continuous on $\mathcal{D}(T^\dagger)$, the density of $\mathcal{D}(T^\dagger)$ in Y ensures that $T^\dagger$ can be extended continuously to Y by

$$\overline{T^\dagger} y := \lim_{n\to\infty} T^\dagger y_n \quad \text{for a sequence } \{y_n\}_{n\in\mathbb{N}} \subset \mathcal{D}(T^\dagger) \text{ with } y_n \to y \in Y.$$

(Since $T^\dagger$ is bounded, it maps Cauchy sequences to Cauchy sequences, and hence $\overline{T^\dagger}$ is well-defined and continuous.) Let now $y \in \overline{\mathcal{R}(T)}$ and $\{y_n\}_{n\in\mathbb{N}} \subset \mathcal{R}(T)$ with $y_n \to y$. As for (3.8), we then have

$$y = P_{\overline{\mathcal{R}}} y = \lim_{n\to\infty} P_{\overline{\mathcal{R}}} y_n = \lim_{n\to\infty} TT^\dagger y_n = T\overline{T^\dagger} y \in \mathcal{R}(T)$$

and hence that $\overline{\mathcal{R}(T)} = \mathcal{R}(T)$. □

Accordingly, the operator equation $Tx = y$ is called *ill-posed in the sense of Nashed* if $\mathcal{R}(T)$ is not closed. Unfortunately, this already excludes many interesting compact operators.

Corollary 3.8. *If $K \in \mathcal{K}(X, Y)$ has infinite-dimensional range $\mathcal{R}(K)$, then $K^\dagger$ is not continuous.*

Proof. Assume to the contrary that $K^\dagger$ is continuous. Then $\mathcal{R}(K)$ is closed by Theorem 3.7, and thus the operator $\tilde{K}$ defined via Theorem 3.3 has a continuous inverse $\tilde{K}^{-1} \in \mathbb{L}(\mathcal{R}(K), \mathcal{N}(K)^\perp)$. Now, K and therefore also $K \circ \tilde{K}^{-1}$ are compact. By

$$K\tilde{K}^{-1} y = y \qquad \text{for all } y \in \mathcal{R}(K),$$

this implies that the identity $\mathrm{Id} : \mathcal{R}(K) \to \mathcal{R}(K)$ is compact as well, which is only possible if $\mathcal{R}(K)$ is finite-dimensional. □

For compact operators, the third condition for well-posedness in the sense of Hadamard therefore has to handled by other methods, which we will study in the following chapters.

3.2 SINGULAR VALUE DECOMPOSITION OF COMPACT OPERATORS

We now characterize the Moore–Penrose inverse of compact operators $K \in \mathcal{K}(X, Y)$ via orthonormal systems. We would like to do this using a spectral decomposition, which however exists only for selfadjoint operators. But by Corollary 3.6, we can equivalently consider the Moore–Penrose inverse of K^*K, which *is* selfadjoint; this leads to the *singular value decomposition.*

Theorem 3.9. *For every $K \in \mathcal{K}(X, Y)$, there exist* *(i)* *a null sequence* $\{\sigma_n\}_{n\in\mathbb{N}}$ *with* $\sigma_1 \geq \sigma_2 \geq \cdots > 0$,

(ii) *an orthonormal basis* $\{u_n\}_{n\in\mathbb{N}} \subset Y$ *of* $\overline{\mathcal{R}(K)}$,

(iii) *an orthonormal basis* $\{v_n\}_{n\in\mathbb{N}} \subset X$ *of* $\overline{\mathcal{R}(K^*)}$

(possibly finite) with

$$Kv_n = \sigma_n u_n \quad \text{and} \quad K^* u_n = \sigma_n v_n \qquad \text{for all } n \in \mathbb{N} \tag{3.9}$$

and

$$Kx = \sum_{n\in\mathbb{N}} \sigma_n \,(x \,|\, v_n)_X \, u_n \qquad \text{for all } x \in X. \tag{3.10}$$

A sequence $\{(\sigma_n, u_n, v_n)\}_{n\in\mathbb{N}}$ *satisfying the* singular value decomposition (3.10) *is called* singular system.

Proof. Since $K^*K : X \to X$ is compact and selfadjoint, the Spectral Theorem 2.5 yields a null sequence $\{\lambda_n\}_{n\in\mathbb{N}} \subset \mathbb{R} \setminus \{0\}$ (ordered by decreasing magnitude) and an orthonormal system $\{v_n\}_{n\in\mathbb{N}} \subset X$ of corresponding eigenvectors with

$$K^*Kx = \sum_{n\in\mathbb{N}} \lambda_n \,(x \,|\, v_n)_X \, v_n \qquad \text{for all } x \in X.$$

Since $\lambda_n = \lambda_n \|v_n\|_X^2 = (\lambda_n v_n \,|\, v_n)_X = (K^*Kv_n \,|\, v_n)_X = \|Kv_n\|_Y^2 > 0$, we can define for all $n \in \mathbb{N}$

$$\sigma_n := \sqrt{\lambda_n} > 0 \qquad \text{and} \qquad u_n := \sigma_n^{-1} K v_n \in Y.$$

The latter form an orthonormal system due to

$$\left(u_i \,\middle|\, u_j\right)_Y = \frac{1}{\sigma_i \sigma_j} \left(Kv_i \,\middle|\, Kv_j\right)_Y = \frac{1}{\sigma_i \sigma_j} \left(K^*Kv_i \,\middle|\, v_j\right)_X = \frac{\lambda_i}{\sigma_i \sigma_j} \left(v_i \,\middle|\, v_j\right)_X = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{else.} \end{cases}$$

Furthermore, we have for all $n \in \mathbb{N}$ that

$$K^* u_n = \sigma_n^{-1} K^* K v_n = \sigma_n^{-1} \lambda_n v_n = \sigma_n v_n.$$

Theorem 2.5 also yields that $\{v_n\}_{n\in\mathbb{N}}$ is an orthonormal basis of $\overline{\mathcal{R}(K^*K)}$. In addition, $\overline{\mathcal{R}(K^*K)} = \overline{\mathcal{R}(K^*)}$, since for any $x \in \mathcal{R}(K^*)$, there exists a sequence $\{y_n\}_{n\in\mathbb{N}} \subset Y$ with $K^* y_n \to x$; in particular, we can take $y_n \in \mathcal{N}(K^*)^\perp = \overline{\mathcal{R}(K)}$, and a diagonal argument shows $x \in \overline{\mathcal{R}(K^*K)}$. (The other direction is obvious.) Hence, $\{v_n\}_{n\in\mathbb{N}}$ is an orthonormal basis of $\overline{\mathcal{R}(K^*)} = \mathcal{N}(K)^\perp$, and therefore

$$Kx = KP_{\mathcal{N}^\perp} x = K\left(\sum_{n\in\mathbb{N}} (x \,|\, v_n)_X \, v_n\right) \qquad \text{for all } x \in X.$$

From this, we obtain the singular value decomposition (3.10) by "pushing" K through the series representation. Since we will repeatedly apply such arguments in the following, we justify this step in detail. First, we set $x_N := \sum_{n=1}^{N} (x \mid v_n)_X \, v_n$ for any $x \in X$ and $N \in \mathbb{N}$. Then we clearly have $x_N \to P_{\mathcal{N}^\perp} x$ as $N \to \infty$ and hence by continuity of K also

$$Kx = K\,(P_{\mathcal{N}^\perp} x) = K(\lim_{N\to\infty} x_N) = \lim_{N\to\infty} Kx_N = \lim_{N\to\infty} \sum_{n=1}^{N} (x \mid v_n)_X \, Kv_n = \sum_{n\in\mathbb{N}} (x \mid v_n)_X \, Kv_n. \tag{3.11}$$

We thus have for all $x \in X$ that

$$Kx = \sum_{n\in\mathbb{N}} (x \mid v_n)_X \, Kv_n = \sum_{n\in\mathbb{N}} (x \mid v_n)_X \, \sigma_n u_n = \sum_{n\in\mathbb{N}} (x \mid K^* u_n)_X \, u_n = \sum_{n\in\mathbb{N}} (Kx \mid u_n)_X \, u_n.$$

The first equation yields (3.10), while the last implies that $\{u_n\}_{n\in\mathbb{N}}$ is an orthonormal basis of $\overline{\mathcal{R}(K)}$. □

Since the eigenvalues of K^*K with eigenvector v_n are exactly the eigenvalues of KK^* with eigenvector u_n, this also yields by (3.9) a singular value decomposition of K^*, i.e.,

$$K^* y = \sum_{n\in\mathbb{N}} \sigma_n \, (y \mid u_n)_Y \, v_n \qquad \text{for all } y \in Y. \tag{3.12}$$

We now use the singular value decomposition of K to characterize the domain $\mathcal{D}(K^\dagger) = \mathcal{R}(K) \oplus \mathcal{R}(K)^\perp$ of the Moore–Penrose inverse $K^\dagger$. As was already discussed before Theorem 3.7, this reduces to the question whether $y \in \overline{\mathcal{R}(K)}$ is in fact an element of $\mathcal{R}(K)$.

Theorem 3.10. *Let $K \in \mathcal{K}(X, Y)$ with singular system $\{(\sigma_n, u_n, v_n)\}_{n\in\mathbb{N}}$ and $y \in \overline{\mathcal{R}(K)}$. Then $y \in \mathcal{R}(K)$ if and only if the* Picard condition

$$\sum_{n\in\mathbb{N}} \sigma_n^{-2} |\,(y \mid u_n)_Y\,|^2 < \infty \tag{3.13}$$

is satisfied. In this case,

$$K^\dagger y = \sum_{n\in\mathbb{N}} \sigma_n^{-1} \, (y \mid u_n)_Y \, v_n. \tag{3.14}$$

Proof. Let $y \in \mathcal{R}(K)$, i.e., there exists $x \in X$ with $Kx = y$. Then

$$(y \mid u_n)_Y = (x \mid K^* u_n)_X = \sigma_n \, (x \mid v_n)_X \qquad \text{for all } n \in \mathbb{N},$$

and the Bessel inequality (2.2) yields

$$\sum_{n\in\mathbb{N}} \sigma_n^{-2} |\,(y \mid u_n)_Y\,|^2 = \sum_{n\in\mathbb{N}} |\,(x \mid v_n)_X\,|^2 \le \|x\|_X^2 < \infty.$$

Conversely, let $y \in \overline{\mathcal{R}(K)}$ satisfy (3.13), which implies that $\{\sum_{n=1}^N \sigma_n^{-2} |(y \mid u_n)_Y|^2\}_{N\in\mathbb{N}}$ is a Cauchy sequence. Then $\{x_N\}_{N\in\mathbb{N}}$ defined by

$$x_N := \sum_{n=1}^{N} \sigma_n^{-1} (y \mid u_n)_Y v_n$$

is a Cauchy sequence as well, since $\{v_n\}_{n\in\mathbb{N}}$ forms an orthonormal system and thus

$$\|x_N - x_M\|_X^2 = \|\textstyle\sum_{n=N+1}^{M} \sigma_n^{-1} (y \mid u_n)_Y v_n\|_X^2 = \displaystyle\sum_{n=N+1}^{M} |\sigma_n^{-1} (y \mid u_n)_Y|^2 \to 0 \quad \text{as } N, M \to \infty.$$

Furthermore, $\{v_n\}_{n\in\mathbb{N}} \subset \overline{\mathcal{R}(K^*)}$. Hence, $\{x_N\}_{N\in\mathbb{N}} \subset \overline{\mathcal{R}(K^*)}$ converges to some

$$x := \sum_{n\in\mathbb{N}} \sigma_n^{-1} (y \mid u_n)_Y v_n \in \overline{\mathcal{R}(K^*)} = \mathcal{N}(K)^\perp$$

by the closedness of $\overline{\mathcal{R}(K^*)}$. Now we have as in (3.11) that

$$Kx = \sum_{n\in\mathbb{N}} \sigma_n^{-1} (y \mid u_n)_Y K v_n = \sum_{n\in\mathbb{N}} (y \mid u_n)_Y u_n = P_{\overline{\mathcal{R}}} y = y,$$

which implies that $y \in \mathcal{R}(K)$.

Finally, $Kx = P_{\overline{\mathcal{R}}} y$ for $x \in \mathcal{N}(K)^\perp$ is equivalent to $x = K^\dagger y$ by Theorem 3.5, which also shows (3.14). □

The Picard condition states that a minimum norm solution can only exist if the "Fourier coefficients" $(y \mid u_n)_Y$ of y decay fast enough compared to the singular values σ_n. The representation (3.14) also shows how perturbations of y relate to perturbations of $x^\dagger$: If $y^\delta = y + \delta u_n$ for some $\delta > 0$ and $n \in \mathbb{N}$, then

$$\|K^\dagger y^\delta - K^\dagger y\|_X = \delta \|K^\dagger u_n\|_X = \sigma_n^{-1}\delta \to \infty \quad \text{as } n \to \infty,$$

and the faster the singular values decay, the more the error is amplified for given n. Hence one distinguishes

- *moderately ill-posed* problems, for which there exist $c, r > 0$ with $\sigma_n \geq c n^{-r}$ for all $n \in \mathbb{N}$ (i.e., σ_n decays at most polynomially), and
- *severely ill-posed* problems, for which this is not the case. If $\sigma_n \leq c e^{-n^r}$ for all $n \in \mathbb{N}$ and $c, r > 0$ (i.e., σ_n decays at least exponentially), the problem is called *exponentially ill-posed*.

For exponentially ill-posed problems, one can in general not expect to obtain a solution that is more than a very rough approximation. On the other hand, if $\mathcal{R}(K)$ finite-dimensional, then the sequence $\{\sigma_n\}_{n\in\mathbb{N}}$ is finite and the error stays bounded; in this case, $K^\dagger$ is continuous as expected.

The singular value decomposition is a valuable analytical tool, but its explicit computation for a concrete operator is in general quite involved. We again consider differentiation as an elementary example.

Example 3.11. Let $X = L^2(\Omega)$ for $\Omega = (0,1)$ and let $K \in \mathcal{K}(X,X)$ be an integral operator defined via

$$[Kx](t) = \int_0^1 k(s,t)x(s)\,ds \qquad \text{with} \qquad k(s,t) = \begin{cases} 1 & \text{if } s \leq t, \\ 0 & \text{else.} \end{cases}$$

If $x = y'$ for some $y \in C^1([0,1])$ with $y(0) = 0$, then

$$[Kx](t) = \int_0^t x(s)\,ds = y(t) - y(0) = y(t),$$

i.e., the derivative y of $y \in C^1([0,1])$ is a solution of the operator equation $Kx = y$ (which is also meaningful for $y \in L^2(\Omega)$ but may not admit a solution then).

The corresponding adjoint operator is given by

$$[K^*y](t) = \int_0^1 k(t,s)y(s)\,ds = \int_t^1 y(s)\,ds,$$

since $k(t,s) = 1$ for $s \geq t$ and 0 else. We now compute the eigenvalues and eigenvectors of K^*K, i.e., any $\lambda > 0$ and $v \in L^2(\Omega)$ with

$$\lambda v(t) = [K^*Kv](t) = \int_t^1 \int_0^s v(r)\,dr\,ds. \tag{3.15}$$

We first proceed formally. Inserting $t = 1$ yields $\lambda v(1) = 0$ and therefore $v(1) = 0$. Differentiating (3.15) yields

$$\lambda v'(t) = \frac{d}{dt}\left(-\int_1^t \int_0^s v(r)\,dr\,ds\right) = -\int_0^t v(r)\,dr,$$

which for $t = 0$ implies that $v'(0) = 0$. Differentiating again now leads to the ordinary differential equation

$$\lambda v''(t) + v(t) = 0$$

which has the general solution

$$v(t) = c_1 \sin(\sigma^{-1}t) + c_2 \cos(\sigma^{-1}t)$$

for $\sigma := \sqrt{\lambda}$ and constants c_1, c_2 that have yet to be determined. For this, we insert the boundary conditions $v'(0) = 0$ and $v(1) = 0$, which yields $c_1 = 0$ and $c_2 \cos(\sigma^{-1}) = 0$, respectively. Since $c_2 = 0$ leads to the trivial solution $v = 0$ and eigenvectors are by definition not trivial, we must have $\cos(\sigma^{-1}) = 0$; the only candidates for the singular values σ_n are therefore the reciprocal roots of the cosine, i.e.,

$$\sigma_n = \frac{2}{(2n-1)\pi}, \qquad n \in \mathbb{N}.$$

From this, we obtain the eigenvectors

$$v_n(t) = \sqrt{2}\cos\left((n-\tfrac{1}{2})\pi\, t\right), \qquad n \in \mathbb{N},$$

where the constant $c_2 = \sqrt{2}$ is chosen such that $\|v_n\|_{L^2(\Omega)} = 1$. We further compute

$$u_n := \sigma_n^{-1} K v_n = (n-\tfrac{1}{2})\pi \int_0^t \sqrt{2}\cos\left((n-\tfrac{1}{2})\pi\, s\right) ds = \sqrt{2}\sin\left((n-\tfrac{1}{2})\pi\, t\right), \quad n \in \mathbb{N}.$$

Now we have $v_n, u_n \in L^2(\Omega)$, and it is straightforward to verify that σ_n^2 and v_n satisfy the eigenvalue relation (3.15). As in the proof of Theorem 3.9, this yields a singular value decomposition of K and thus a singular system $\{(\sigma_n, u_n, v_n)\}_{n\in\mathbb{N}}$.

Since $\sigma_n = O(\frac{1}{n})$, this implies that differentiation (in this formulation) is a moderately ill-posed problem. Furthermore, the Picard condition (3.13) for $y \in L^2(\Omega)$ is given by

$$\sum_{n\in\mathbb{N}} \frac{1}{4}(2n-1)^2\pi^2 |\,(y \mid u_n)_{L^2}\,|^2 < \infty.$$

It is now possible to show that $\{u_n\}_{n\in\mathbb{N}}$ is an orthonormal basis of $L^2(\Omega)$ (which are not unique) and thus satisfy

$$y = \sum_{n\in\mathbb{N}} (y \mid u_n)_{L^2}\, u_n.$$

Formally differentiating the Fourier series term by term then yields

$$z := \sum_{n\in\mathbb{N}} (y \mid u_n)_{L^2}\, u_n' = \sum_{n\in\mathbb{N}} \left(n - \tfrac{1}{2}\right) \pi\, (y \mid u_n)_{L^2}\, v_n.$$

The Picard condition is thus equivalent to the condition that $\|z\|_X^2 < \infty$ and hence that the formally differentiated series converges (in $L^2(\Omega)$); in this case $K^\dagger y = z$. If y is continuously differentiable, this convergence is even uniform and we obtain that $y' = z = K^\dagger y$.

The singular value decomposition allows defining functions of compact operators, which will be a fundamental tool in the following chapters. Let $\varphi : [0,\infty) \to \mathbb{R}$ be a piecewise continuous and locally bounded function. We then define for $K \in \mathcal{K}(X,Y)$ with singular system $\{(\sigma_n, u_n, v_n)\}_{n\in\mathbb{N}}$ the operator $\varphi(K^*K) : X \to X$ by

$$\varphi(K^*K)x = \sum_{n\in\mathbb{N}} \varphi(\sigma_n^2)\,(x\,|\,v_n)_X\, v_n + \varphi(0)P_{\mathcal{N}(K)}x \qquad \text{for all } x \in X. \tag{3.16}$$

This series converges in X since φ is only evaluated on the closed and bounded interval $[0, \sigma_1^2] = [0, \|K\|^2_{\mathbb{L}(X,Y)}]$. Furthermore, the Bessel inequality implies that

$$\|\varphi(K^*K)\|_{\mathbb{L}(X,X)} \le \sup_{n\in\mathbb{N}} |\varphi(\sigma_n^2)| + \varphi(0) \le 2 \sup_{\lambda\in[0,\|K\|^2_{\mathbb{L}(X,Y)}]} |\varphi(\lambda)| < \infty, \tag{3.17}$$

i.e., $\varphi(K^*K) \in \mathbb{L}(X,X)$.

In particular, we consider here power functions $\varphi(t) = t^r$ for $r \ge 0$ and especially the following examples.

Example 3.12. Let $K \in \mathcal{K}(X,Y)$.

(i) For $\varphi(t) = 1$ we have $\varphi(K^*K) = \mathrm{Id}$ since for all $x \in X$,

$$\varphi(K^*K)x = \sum_{n\in\mathbb{N}} (x\,|\,v_n)_X\, v_n + P_{\mathcal{N}(K)}x = P_{\overline{\mathcal{R}(K^*)}}x + P_{\mathcal{N}(K)}x = x$$

due to $\overline{\mathcal{R}(K^*)} = \mathcal{N}(K)^\perp$.

(ii) For $\varphi(t) = t$ we have $\varphi(K^*K) = K^*K$ due to $\varphi(0) = 0$ and the spectral theorem.

(iii) For $\varphi(t) = \sqrt{t}$ we call $|K| := \varphi(K^*K)$ the *absolute value* of K; since $\sigma_n > 0$, we have

$$|K|x = \sum_{n\in\mathbb{N}} \sigma_n\,(x\,|\,v_n)_X\, v_n \qquad \text{for all } x \in X.$$

Comparing Example 3.12 (iii) with the singular value decomposition (3.10) shows that $|K|$ essentially has the same behavior as K, the only difference being that the former maps to X instead of Y. This is illustrated by the following properties, which will be used later.

Lemma 3.13. *Let $K \in \mathcal{K}(X,Y)$. Then*

(i) $|K|^{r+s} = |K|^r \circ |K|^s$ *for all* $r, s \ge 0$*;*

(ii) $|K|^r$ *is selfadjoint for all* $r \ge 0$*;*

(iii) $\||K|x\|_X = \|Kx\|_Y$ *for all* $x \in X$*;*

(iv) $\mathcal{R}(|K|) = \mathcal{R}(K^*)$.

Proof. Ad (i): This follows directly from

$$\begin{aligned}|K|^{r+s}x &= \sum_{n\in\mathbb{N}} \sigma_n^{r+s} (x \mid v_n)_X v_n = \sum_{n\in\mathbb{N}} \sigma_n^{r} \left(\sigma_n^{s} (x \mid v_n)_X\right) v_n \\ &= \sum_{n\in\mathbb{N}} \sigma_n^{r} \left(\sum_{m\in\mathbb{N}} \sigma_m^{s} (x \mid v_m)_X v_m \,\middle|\, v_n\right)_X v_n \\ &= \sum_{n\in\mathbb{N}} \sigma_n^{r} (|K|^s x \mid v_n)_X v_n = |K|^r(|K|^s x)\end{aligned}$$

since $\{v_n\}_{n\in\mathbb{N}}$ is an orthonormal system.

Ad (ii): For any $x, z \in X$ and $r \geq 0$, the bilinearity and symmetry of the inner product implies that

$$(|K|^r x \mid z)_X = \sum_{n\in\mathbb{N}} \sigma_n^{r} (x \mid v_n)_X (v_n \mid z)_X = (x \mid |K|^r z)_X .$$

Ad (iii): This follows from (i), (ii), and

$$\| |K| x\|_X^2 = (|K|x \mid |K|x)_X = \left(|K|^2 x \,\middle|\, x\right)_X = (K^*Kx \mid x)_X = (Kx \mid Kx)_X = \|Kx\|_X^2.$$

Ad (iv): Let $\{(\sigma_n, u_n, v_n)\}_{n\in\mathbb{N}}$ be a singular system of K. Then $\{(\sigma_n, v_n, u_n)\}_{n\in\mathbb{N}}$ is a singular system of K^*, and – by definition – $\{(\sigma_n, v_n, v_n)\}_{n\in\mathbb{N}}$ is a singular system of $|K|$. Now $x \in \mathcal{R}(K^*)$ if and only if $Kx \in \mathcal{R}(KK^*)$ and $x \in \mathcal{N}(K)^\perp$. The Picard condition for $Kx \in \mathcal{R}(KK^*)$ is

$$\infty > \sum_{n\in\mathbb{N}} \sigma_n^{-4} |\,(Kx \mid u_n)_Y\,|^2 = \sum_{n\in\mathbb{N}} \sigma_n^{-4} |\,(x \mid K^* u_n)_X\,|^2 = \sum_{n\in\mathbb{N}} \sigma_n^{-2} |\,(x \mid v_n)_X\,|^2.$$

But this is also the Picard condition for $x \in \mathcal{R}(|K|)$ (compare the proof of Theorem 3.10), which for $x \in \mathcal{N}(K)^\perp$ is even a necessary condition. □

The proof of Lemma 3.13 (iv) already indicates that we can use $|K|$ to formulate a variant of the Picard condition for $x \in \overline{\mathcal{R}(K^*)}$ (instead of $y \in \overline{\mathcal{R}(K)}$); we will use this in a following chapter to characterize minimum norm solutions that can be particularly well approximated.

We finally need the following inequality.

Lemma 3.14. *Let $K \in \mathcal{K}(X, Y)$. Then any $r > s \geq 0$ and $x \in X$ satisfy the* interpolation inequality

$$\| |K|^s x\|_X \leq \| |K|^r x\|_X^{\frac{s}{r}} \, \|x\|_X^{1-\frac{s}{r}}. \tag{3.18}$$

Proof. By definition of $|K|^s$,

$$\||K|^s x\|_X^2 = \sum_{n\in\mathbb{N}} \sigma_n^{2s} |(x, v_n)_X|^2, \tag{3.19}$$

which together with the Bessel inequality immediately yields the claim for $s = 0$.

For $s > 0$, we apply the Hölder inequality

$$\sum_{n\in\mathbb{N}} a_n b_n \leq \left(\sum_{n\in\mathbb{N}} a_n^p\right)^{\frac{1}{p}} \left(\sum_{n\in\mathbb{N}} b_n^q\right)^{\frac{1}{q}} \qquad \text{for} \quad \frac{1}{p} + \frac{1}{q} = 1$$

to

$$a_n := \sigma_n^{2s} |\left(x \mid v_n\right)_X|^{2\frac{s}{r}}, \qquad b_n := |\left(x \mid v_n\right)_X|^{2-2\frac{s}{r}}, \qquad p = \frac{r}{s}, \qquad q = \frac{r}{r-s}.$$

Then, (3.19) and the Bessel inequality yield

$$\begin{aligned} \||K|^s x\|_X^2 &\leq \left(\sum_{n\in\mathbb{N}} \sigma_n^{2r} |\left(x \mid v_n\right)_X|^2\right)^{\frac{s}{r}} \left(\sum_{n\in\mathbb{N}} |\left(x \mid v_n\right)_X|^2\right)^{1-\frac{s}{r}} \\ &\leq \||K|^r x\|_X^{2\frac{s}{r}} \|x\|_X^{2(1-\frac{s}{r})}, \end{aligned}$$

and the claim follows after taking the square root. □

4 REGULARIZATION METHODS

As shown in the last chapter, the ill-posed operator equation $Tx = y$ admits for any $y \in \mathcal{D}(T^\dagger)$ a unique minimum norm solution $x^\dagger = T^\dagger y$. In practice, one however usually does not have access to the "exact data" y but only to a "noisy measurement" $y^\delta \in B_\delta(y)$, i.e., satisfying

$$\|y - y^\delta\|_Y \leq \delta,$$

where $\delta > 0$ is the *noise level.* Since $T^\dagger$ is not continuous in general, $T^\dagger y^\delta$ is not guaranteed to be a good approximation of $x^\dagger$ even for $y^\delta \in \mathcal{D}(T^\dagger)$. The goal is therefore to construct an approximation x_α^δ that on the one hand depends continuously on y^δ – and thus on δ – and on the other hand can through the choice of a *regularization parameter* $\alpha > 0$ be brought as close to $x^\dagger$ as the noise level δ allows. In particular, for $\delta \to 0$ and an appropriate choice of $\alpha(\delta)$, we want to ensure that $x_{\alpha(\delta)}^\delta \to x^\dagger$. A method which constructs such an approximation is called *regularization method.*

4.1 REGULARIZATION AND PARAMETER CHOICE

For linear operators between Hilbert spaces, such constructions can be defined through *regularization operators*, which can be considered as a continuous replacement for the unbounded pseudoinverse $T^\dagger$. This leads to the following definition.

Definition 4.1. Let $T \in \mathbb{L}(X, Y)$ be a bounded linear operator between the Hilbert spaces X and Y. A family $\{R_\alpha\}_{\alpha>0}$ of linear operators $R_\alpha : Y \to X$ is called a *regularization* (of $T^\dagger$) if

(i) $R_\alpha \in \mathbb{L}(Y, X)$ for all $\alpha > 0$;

(ii) $R_\alpha y \to T^\dagger y$ as $\alpha \to 0$ and all $y \in \mathcal{D}(T^\dagger)$.

A regularization is therefore a pointwise approximation of the Moore–Penrose inverse by continuous operators. However, the Banach–Steinhaus Theorem implies that the convergence cannot be unifom if $T^\dagger$ is not continuous.

Theorem 4.2. *Let $T \in \mathbb{L}(X, Y)$ and $\{R_\alpha\}_{\alpha>0} \subset \mathbb{L}(Y, X)$ be a regularization. If $T^\dagger$ is not continuous, then $\{R_\alpha\}_{\alpha>0}$ is not uniformly bounded. In particular, then there exists a $y \in Y$ and a null sequence $\{\alpha_n\}_{n\in\mathbb{N}}$ with $\|R_{\alpha_n} y\|_X \to \infty$.*

Proof. Assume to the contrary that no such $y \in Y$ exists. Then the family $\{R_\alpha\}_{\alpha>0} \subset \mathbb{L}(Y,X)$ is bounded pointwise and hence uniformly by the Banach–Steinhaus Theorem 1.5. Thus there exists an $M > 0$ with $\|R_\alpha\|_{\mathbb{L}(Y,X)} \leq M$ for all $\alpha > 0$. Together with the pointwise convergence $R_\alpha \to T^\dagger$ on the dense subset $\mathcal{D}(T^\dagger) \subset Y$, Corollary 1.6 yields convergence on all of $\overline{\mathcal{D}(T^\dagger)} = Y$. By Corollary 1.7, $T^\dagger$ is then continuous, and the claim follows by contraposition. □

In fact, under an additional assumption, $R_{\alpha_n} y$ has to diverge for *all* $y \notin \mathcal{D}(T^\dagger)$.

Theorem 4.3. *Let $T \in \mathbb{L}(X,Y)$ be such that $T^\dagger$ is not continuous, and let $\{R_\alpha\}_{\alpha>0} \subset \mathbb{L}(Y,X)$ be a regularization of $T^\dagger$. If*

$$\sup_{\alpha>0} \|TR_\alpha\|_{\mathbb{L}(Y,Y)} < \infty, \tag{4.1}$$

then $\|R_\alpha y\|_X \to \infty$ as $\alpha \to 0$ and all $y \notin \mathcal{D}(T^\dagger)$.

Proof. Let $y \in Y$ be arbitrary and assume that there exists a null sequence $\{\alpha_n\}_{n\in\mathbb{N}}$ for which $\{R_{\alpha_n} y\}_{n\in\mathbb{N}}$ is bounded. Then there exists a subsequence $\{x_k\}_{k\in\mathbb{N}}$, $x_k := R_{\alpha_{n_k}} y$, with $x_k \rightharpoonup x \in X$. Since bounded linear operators are weakly continuous, this also yields that $Tx_k \rightharpoonup Tx$.

On the other hand, the continuity of T and the pointwise convergence $R_\alpha \to T^\dagger$ on $\mathcal{D}(T^\dagger)$ imply together with Lemma 3.4 (iv) that $TR_\alpha y \to TT^\dagger y = P_{\overline{\mathcal{R}}} y$ for all $y \in \mathcal{D}(T^\dagger)$. The assumption (4.1) and Corollary 1.6 then yield the pointwise convergence of $TR_{\alpha_n} \to P_{\overline{\mathcal{R}}}$ on all of Y. From $Tx_k = TR_{\alpha_{n_k}} y \to P_{\overline{\mathcal{R}}} y$ and $Tx_k \rightharpoonup Tx$, it now follows by the uniqueness of the limit that $Tx = P_{\overline{\mathcal{R}}} y$. Hence $P_{\overline{\mathcal{R}}} y \in \mathcal{R}(T)$ and therefore $y \in \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp = \mathcal{D}(T^\dagger)$, and the claim follows by contraposition. □

However, we can in general not assume that a given noisy measurement $y^\delta \in B_\delta(y)$ is an element of $\mathcal{D}(T^\dagger)$. We therefore have to consider the *regularization error*

$$\begin{aligned} \|R_\alpha y^\delta - T^\dagger y\|_X &\leq \|R_\alpha y^\delta - R_\alpha y\|_X + \|R_\alpha y - T^\dagger y\|_X \\ &\leq \delta \|R_\alpha\|_{\mathbb{L}(Y,X)} + \|R_\alpha y - T^\dagger y\|_X. \end{aligned} \tag{4.2}$$

This decomposition is a fundamental tool of regularization theory, and we will meet it repeatedly throughout the following. Here the first term describes the *(propagated) data error*, which by Theorem 4.2 cannot be bounded for $\alpha \to 0$ as long as $\delta > 0$. The second term describes the *approximation error*, which due to the assumed pointwise convergence for $\alpha \to 0$ does tend to zero. To obtain a reasonable approximation, we thus have to choose α in a suitable dependence of δ such that the total regularization error vanishes as $\delta \to 0$.

Definition 4.4. A function $\alpha : \mathbb{R}^+ \times Y \to \mathbb{R}^+$, $(\delta, y^\delta) \mapsto \alpha(\delta, y^\delta)$, is called a *parameter choice rule*. We distinguish

(i) *a priori choice rules* that only depend on δ;

(ii) *a posteriori choice rules* that depend on δ and y^δ;

(iii) *heuristic choice rules* that only depend on y^δ.

If $\{R_\alpha\}_{\alpha>0}$ is a regularization of $T^\dagger$ and α is a parameter choice rule, the pair (R_α, α) is called a *(convergent) regularization method* if

$$\lim_{\delta\to 0} \sup_{y^\delta \in B_\delta(y)} \|R_{\alpha(\delta,y^\delta)} y^\delta - T^\dagger y\|_X = 0 \qquad \text{for all } y \in \mathcal{D}(T^\dagger). \tag{4.3}$$

We thus demand that the regularization error vanishes for *all* noisy measurements y^δ that are compatible with the noise level $\delta \to 0$.

A PRIORI CHOICE RULES

We first show that every regularization admits an a priori choice rule and hence leads to a convergent regularization method.

Theorem 4.5. *Let $\{R_\alpha\}_{\alpha>0}$ be a regularization of $T^\dagger$. Then there exists an a priori choice rule α such that (R_α, α) is a regularization method.*

Proof. Let $y \in \mathcal{D}(T^\dagger)$ be arbitrary. Since $R_\alpha \to T^\dagger$ pointwise by assumption, there exists for all $\varepsilon > 0$ a $\sigma(\varepsilon) > 0$ such that

$$\|R_{\sigma(\varepsilon)} y - T^\dagger y\|_X \le \frac{\varepsilon}{2}.$$

This defines a monotonically increasing function $\sigma : \mathbb{R}^+ \to \mathbb{R}^+$ with $\lim_{\varepsilon\to 0} \sigma(\varepsilon) = 0$. Similarly, the operator $R_{\sigma(\varepsilon)}$ is continuous for every fixed $\varepsilon > 0$ and hence there exists a $\rho(\varepsilon) > 0$ with

$$\|R_{\sigma(\varepsilon)} z - R_{\sigma(\varepsilon)} y\|_X \le \frac{\varepsilon}{2} \qquad \text{for all } z \in Y \text{ with } \|z - y\|_Y \le \rho(\varepsilon).$$

Again, this defines a function $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ with $\lim_{\varepsilon\to 0} \rho(\varepsilon) = 0$, where we can assume without loss of generality that ρ is strictly increasing and continuous (by choosing $\rho(\varepsilon)$ maximally in case it is not unique). The Inverse Function Theorem thus ensures that there exists a strictly monotone and continuous inverse function ρ^{-1} on $\mathcal{R}(\rho)$ with $\lim_{\delta\to 0} \rho^{-1}(\delta) = 0$. We extend this function monotonically and continuously to $\mathbb{R}^+$ and define our a priori choice rule

$$\alpha : \mathbb{R}^+ \to \mathbb{R}^+, \qquad \delta \mapsto \sigma(\rho^{-1}(\delta)).$$

Then we have in particular $\lim_{\delta\to 0}\alpha(\delta) = 0$. Furthermore, for all $\varepsilon > 0$ there exists a $\delta := \rho(\varepsilon) > 0$ such that $\alpha(\delta) = \sigma(\varepsilon)$ and hence

$$\|R_{\alpha(\delta)}y^\delta - T^\dagger y\|_X \le \|R_{\sigma(\varepsilon)}y^\delta - R_{\sigma(\varepsilon)}y\|_X + \|R_{\sigma(\varepsilon)}y - T^\dagger y\|_X \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

for all $y^\delta \in B_\delta(y)$. This implies that $\|R_{\alpha(\delta)}y^\delta - T^\dagger y\|_X \to 0$ as $\delta \to 0$ for any family $\{y^\delta\}_{\delta>0} \subset B_\delta(y)$. Hence (R_α, α) is a convergent regularization method. □

We can even give a full characterization of a priori choice rules that lead to convergent regularization methods.

Theorem 4.6. *Let $T^\dagger$ no be continuous, $\{R_\alpha\}_{\alpha>0}$ be a regularization, and $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ an a priori choice rule. Then (R_α, α) is a regularization method if and only if*

(i) $\lim_{\delta\to 0} \alpha(\delta) = 0$,

(ii) $\lim_{\delta\to 0} \delta\|R_{\alpha(\delta)}\|_{\mathbb{L}(Y,X)} = 0$.

Proof. The decomposition (4.2) of the regularization error immediately implies that

$$\|R_{\alpha(\delta)}y^\delta - T^\dagger y\|_X \le \delta\|R_{\alpha(\delta)}\|_{\mathbb{L}(Y,X)} + \|R_{\alpha(\delta)}y - T^\dagger y\|_X \to 0 \qquad \text{for } \delta \to 0$$

since the first term vanishes by assumption (ii), while the second vanishes due to the pointwise convergence of regularization operators together with assumption (i).

Conversely, assume that either (i) or (ii) does not hold. If (i) is violated, then $R_{\alpha(\delta)}$ does not converge pointwise to $T^\dagger y$. Hence, (4.3) cannot hold for the constant sequence $y^\delta \equiv y$ and $\delta \to 0$, and therefore (R_α, α) is not a regularization method. If now (i) holds but (ii) is violated, there exists a null sequence $\{\delta_n\}_{n\in\mathbb{N}}$ with $\delta_n\|R_{\alpha(\delta_n)}\|_{\mathbb{L}(Y,X)} \ge C > 0$. We can therefore find a sequence $\{z_n\}_{n\in\mathbb{N}} \subset Y$ with $\|z_n\|_Y = 1$ and $\delta_n\|R_{\alpha(\delta_n)}z_n\|_X \ge C$. Let now $y \in \mathcal{D}(T^\dagger)$ be arbitrary and set $y_n := y + \delta_n z_n$. Then $y_n \in B_{\delta_n}(y)$, but

$$R_{\alpha(\delta_n)}y_n - T^\dagger y = (R_{\alpha(\delta_n)}y - T^\dagger y) + \delta_n R_{\alpha(\delta_n)}z_n \not\to 0$$

since the first term on the right-hand side is a null sequence by (i) and the pointwise convergence of R_α, but the second term is not a null sequence by construction. Hence, (4.3) is violated and (R_α, α) therefore not a regularization method. The claim now follows by contraposition. □

Since $\|R_\alpha\|_{\mathbb{L}(Y,X)} \to \infty$ as $\alpha \to 0$, assumption (ii) states that α cannot tend to zero too fast compared to δ. An a priori choice rule thus usually has the form $\alpha(\delta) = \delta^r$ for some $r \in (0, 1)$.

A POSTERIORI CHOICE RULES

As we will see later, the optimal choice of $\alpha(\delta)$ requires information about the exact (minimum norm) solution $x^\dagger$ that is not easily accessible. Such information is not required for a posteriori choice rules. The main idea behind these is the following: Let again $y \in \mathcal{D}(T^\dagger)$ and $y^\delta \in B_\delta(y)$ and consider the *residual*

$$\|TR_\alpha y^\delta - y^\delta\|_Y.$$

If now $y \in \mathcal{R}(T)$ and $\|y - y^\delta\|_Y = \delta$, even the (desired) minimum norm solution $x^\dagger$ satisfies due to $Tx^\dagger = y$ only

$$\|Tx^\dagger - y^\delta\|_Y = \|y - y^\delta\|_Y = \delta.$$

It is therefore not reasonable to try to obtain a smaller residual for the regularization $R_\alpha y^\delta$ either. This motivates the *Morozov discrepancy principle*: For given $\delta > 0$ and $y^\delta \in B_\delta(y)$ choose $\alpha = \alpha(\delta, y^\delta)$ (as large as possible) such that

$$\|TR_\alpha y^\delta - y^\delta\|_Y \le \tau\delta \qquad \text{for some } \tau > 1 \text{ independent of } \delta \text{ and } y^\delta. \tag{4.4}$$

However, this principle may not be satisfiable: If $y \in \mathcal{R}(T)^\perp \setminus \{0\}$, then even the exact data $y^\delta = y$ and the minimum norm solution $x^\dagger$ only satisfy

$$\|Tx^\dagger - y\|_Y = \|TT^\dagger y - y\|_Y = \|P_{\overline{\mathcal{R}}} y - y\|_Y = \|y\|_Y > \tau\delta$$

for some fixed $\tau > 1$ and δ small enough. We therefore have to assume that this situation cannot occur; for this it is sufficient that $\mathcal{R}(T)$ is dense in Y (since in this case $\mathcal{R}(T)^\perp = \overline{\mathcal{R}(T)}^\perp = \{0\}$).

The practical realization usually consists in choosing a null sequence $\{\alpha_n\}_{n\in\mathbb{N}}$, computing successively $R_{\alpha_n} y^\delta$ for $n = 1, \dots,$ and stopping as soon as the discrepancy principle (4.4) is satisfied for an α_{n^*}. The following theorem justifies this procedure.

Theorem 4.7. *Let $\{R_\alpha\}_{\alpha>0}$ be a regularization of $T^\dagger$ with $\mathcal{R}(T)$ dense in Y, $\{\alpha_n\}_{n\in\mathbb{N}}$ be a strictly decreasing null sequence, and $\tau > 1$. If the family $\{TR_\alpha\}_{\alpha>0}$ is uniformly bounded, then for all $y \in \mathcal{D}(T^\dagger)$, $\delta > 0$ and $y^\delta \in B_\delta(y)$ there exists an $n^* \in \mathbb{N}$ such that*

$$\|TR_{\alpha_{n^*}} y^\delta - y^\delta\|_Y \le \tau\delta < \|TR_{\alpha_n} y^\delta - y^\delta\|_Y \qquad \text{for all } n < n^*. \tag{4.5}$$

Proof. We proceed as in the proof of Theorem 4.3. The family $\{TR_\alpha\}_{\alpha>0}$ converges pointwise to $TT^\dagger = P_{\overline{\mathcal{R}}}$ on $\mathcal{D}(T^\dagger)$ and hence, due to the uniform boundedness, on all of $Y = \overline{\mathcal{D}(T^\dagger)}$. This implies that for all $y \in \mathcal{D}(T^\dagger) = \mathcal{R}(T)$ and $y^\delta \in B_\delta(y)$,

$$\lim_{n\to\infty} \|TR_{\alpha_n} y^\delta - y^\delta\|_Y = \|P_{\overline{\mathcal{R}}} y^\delta - y^\delta\|_Y = 0$$

since $\overline{\mathcal{R}(T)} = Y$. From this, the claim follows. ◻

To show that the discrepancy principle indeed leads to a regularization method, it has to be considered in combination with a concrete regularization. We will do so in the following chapters.

HEURISTIC CHOICE RULES

Heuristic choice rules do not need knowledge of the noise level δ, which is often relevant in practice where this knowledge is not available (sufficiently exactly). However, the following pivotal result – known in the literature as the *Bakushinskiĭ veto*, see [Bakushinskiĭ 1985] – states that this is not possible in general.

Theorem 4.8. *Let $\{R_\alpha\}_{\alpha>0}$ be a regularization of $T^\dagger$. If there exists a heuristic choice rule α such that (R_α, α) is a regularization method, then $T^\dagger$ is continuous.*

Proof. Assuming to the contrary that such a parameter choice rule $\alpha : Y \to \mathbb{R}^+$ exists, we can define the mapping

$$R : Y \to X, \qquad y \mapsto R_{\alpha(y)}y$$

Let nwo $y \in \mathcal{D}(T^\dagger)$ be arbitrary and consider any sequence $\{y_n\}_{n\in\mathbb{N}} \subset \mathcal{D}(T^\dagger)$ with $y_n \to y$. On the one hand, then naturally $y_n \in B_\delta(y_n)$ for all $\delta > 0$ and $n \in \mathbb{N}$, and the assumption (4.3) for fixed $y^\delta = y = y_n$ and $\delta \to 0$ yields that $Ry_n = T^\dagger y_n$ for all $n \in \mathbb{N}$. On the other hand, for $\delta_n := \|y_n - y\|_Y$ we also have $y_n \in B_{\delta_n}(y)$, and in this case passing to the limit $n \to \infty$ in (4.3) shows that

$$T^\dagger y_n = Ry_n = R_{\alpha(y_n)}y_n \to T^\dagger y,$$

i.e., $T^\dagger$ is continuous on $\mathcal{D}(T^\dagger)$. □

In particular for compact operators with infinite-dimensional range, *no* heuristic choice rule can lead to a regularization method. Of course, this does not mean that such methods cannot be used in practice. First, the veto does not rule out choice rules for finite-dimensional ill-posed problems (such as very ill-conditioned linear systems); however, these rules are then by necessity dimension-dependent. Second, a sharp look at the proof shows that the crucial step consists in applying the choice rule to data $y^\delta \in \mathcal{D}(T^\dagger)$. The worst case for the noisy data is therefore $y^\delta \in \mathcal{R}(T)$ (since only this subspace of $\mathcal{D}(T^\dagger)$ plays a role due to $\mathcal{R}(T)^\perp = \mathcal{N}(T^\dagger)$), and in this case convergence cannot be guaranteed. In many interesting cases, however, T is a compact (i.e., smoothing) operator, while errors have a more random character and therefore do not typically lie in $\mathcal{R}(T)$. Heuristic choice rules can therefore indeed work in "usual" situations. In fact, it is possible to show under the additional assumption $y^\delta \notin \mathcal{D}(T^\dagger)$ that a whole class of popular heuristic choice rules lead to a regularization method. Here, too, we need to consider the combination with a concrete regularization operator but already give some examples.

(i) The *quasi-optimality principle* picks a finite strictly decreasing sequence $\{\alpha_n\}_{n\in\{1,\dots,N\}}$ and chooses $\alpha(y^\delta) = \alpha_{n^*}$ as the one satisfying

$$n^* \in \arg\min_{1\le n<N} \|R_{\alpha_{n+1}}y^\delta - R_{\alpha_n}y^\delta\|_X.$$

(ii) The *Hanke–Raus rule* chooses

$$\alpha(y^\delta) \in \arg\min_{\alpha>0} \frac{1}{\sqrt{\alpha}}\|TR_\alpha y^\delta - y^\delta\|_Y.$$

(iii) The *L-curve criterion*[1] chooses

$$\alpha(y^\delta) \in \arg\min_{\alpha>0} \|R_\alpha y^\delta\|_X \|TR_\alpha y^\delta - y^\delta\|_Y.$$

All of these methods in one way or another work by using the residual to obtain a reasonably close approximation of the noise level that is then used similarly as in an a priori or a posteriori choice rules. An extensive numerical comparison of these and other choice rules can be found in [Bauer & Lukas 2011].

4.2 CONVERGENCE RATES

A central goal in the regularization of inverse problems is to obtain error estimates of the form

$$\|R_{\alpha(\delta,y^\delta)} y^\delta - T^\dagger y\|_X \leq \psi(\delta)$$

for an increasing function $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ with $\lim_{t\to 0} \psi(t) = 0$. In particular, we are interested in the *worst-case error*

$$\mathcal{E}(y,\delta) := \sup_{y^\delta \in B_\delta(y)} \|R_{\alpha(\delta,y^\delta)} y^\delta - T^\dagger y\|_X \tag{4.6}$$

(which for regularization methods converges to zero as $\delta \to 0$ and any $y \in \mathcal{D}(T^\dagger)$ by (4.3)). Here, ψ has to depend in some form on y since otherwise it would be possible to give regularization error estimates independently of y and y^δ – but since the convergence of $R_\alpha \to T^\dagger$ is merely pointwise but not uniform, such estimates cannot be expected.

Theorem 4.9. *Let (R_α, α) be a regularization method. If there exists a $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ with $\lim_{t\to 0} \psi(t) = 0$ and*

$$\sup_{y \in \mathcal{D}(T^\dagger)\cap B_Y} \mathcal{E}(y,\delta) \leq \psi(\delta), \tag{4.7}$$

then $T^\dagger$ is continuous.

[1] The name is due to the practical realization: If one plots the curve $\alpha \mapsto (\|TR_\alpha y^\delta - y^\delta\|_Y, \|R_\alpha y^\delta\|_X)$ (or, rather, a finite set of points on it) in a doubly logarithmic scale, it often has – more or less – the form of an "L"; the chosen parameter is then the one lying closest to the "knee" of the L.

Proof. Let $y \in \mathcal{D}(T^\dagger) \cap B_Y$ and $\{y_n\}_{n\in\mathbb{N}} \subset \mathcal{D}(T^\dagger) \cap B_Y$ be a sequence with $y_n \to y$. Setting $\delta_n := \|y - y_n\|_Y \to 0$, we than have for $n \to \infty$ that

$$\begin{aligned}\|T^\dagger y_n - T^\dagger y\|_X &\le \|T^\dagger y_n - R_{\alpha(\delta_n,y_n)} y_n\|_X + \|R_{\alpha(\delta_n,y_n)} y_n - T^\dagger y\|_X \\ &\le \mathcal{E}(y_n, \delta_n) + \mathcal{E}(y, \delta_n) \\ &\le 2\psi(\delta_n) \to 0.\end{aligned}$$

Hence $T^\dagger$ is continuous on $\mathcal{D}(T^\dagger) \cap B_Y$ and thus, by linearity of $T^\dagger$, on all of $\mathcal{D}(T^\dagger)$. □

This implies that the convergence can be arbitrarily slow; knowledge of δ alone is therefore not sufficient to give error estimates – we thus need additional assumptions on the exact data y or, equivalently, the wanted minimum norm solution $x^\dagger = T^\dagger y$. As the proof of Theorem 4.9 shows, the existence of convergence rates is closely tied to the continuity of $T^\dagger$ on closed subsets. We therefore consider for $\mathcal{M} \subset X$ and $\delta > 0$ the quantity

$$\varepsilon(\mathcal{M}, \delta) := \sup \{\|x\|_X \mid x \in \mathcal{M},\ \|Tx\|_Y \le \delta\},$$

which can be interpreted as a *modulus of conditional continuity* of $T^\dagger : \mathcal{R}(T) \cap \delta B_Y \to \mathcal{M}$. This modulus is in fact a lower bound for the worst-case error. Since both $\varepsilon(\mathcal{M}, \delta)$ and $\mathcal{E}(y, \delta)$ are not finite if $\mathcal{M} \cap \mathcal{N}(K) \ne \{0\}$ and $\mathcal{M}$ are unbounded, we will only consider the more interesting case that $\mathcal{M} \subset \mathcal{N}(T)^\perp$.

Theorem 4.10. *Let (R_α, α) be a regularization method. Then for all $\delta > 0$ and $\mathcal{M} \subset \mathcal{N}(T)^\perp$,*

$$\sup_{y \in \mathcal{D}(T^\dagger), T^\dagger y \in \mathcal{M}} \mathcal{E}(y, \delta) \ge \varepsilon(\mathcal{M}, \delta).$$

Proof. Let $x \in \mathcal{M}$ with $\|Tx\|_Y \le \delta$. For $y^\delta = 0$, we then deduce from $x \in \mathcal{N}(T)^\perp$ that

$$\|x\|_X = \|T^\dagger Tx - R_{\alpha(\delta,0)} 0\|_X \le \mathcal{E}(Tx, \delta)$$

and hence

$$\varepsilon(\mathcal{M}, \delta) = \sup_{x \in \mathcal{M}, \|Tx\|_Y \le \delta} \|x\|_X \le \sup_{x \in \mathcal{M}, \|Tx\|_Y \le \delta} \mathcal{E}(Tx, \delta) \le \sup_{T^\dagger y \in \mathcal{M}, y \in \mathcal{D}(T^\dagger)} \mathcal{E}(y, \delta)$$

since $\mathcal{D}(T^\dagger) = \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$ and $\mathcal{R}(T)^\perp = \mathcal{N}(T^\dagger)$. □

For an appropriate choice of $\mathcal{M}$, we can now derive sharp bounds on $\varepsilon(\mathcal{M}, \delta)$. We consider here for compact operators $K \in \mathcal{K}(X, Y)$ subsets of the form

$$X_{\nu,\rho} = \{|K|^\nu w \in X \mid \|w\|_X \le \rho\} \subset \mathcal{R}(|K|^\nu).$$

The definition of $|K|^\nu w$ via the spectral decomposition of K implies in particular that $X_{\nu,\rho} \subset \overline{\mathcal{R}(K^*)} = \mathcal{N}(K)^\perp$.

Theorem 4.11. *Let $K \in \mathcal{K}(X, Y)$ and $\nu, \rho > 0$. Then for all $\delta > 0$,*

$$\varepsilon(X_{\nu,\rho}, \delta) \leq \delta^{\frac{\nu}{\nu+1}} \rho^{\frac{1}{\nu+1}}.$$

Proof. Let $x \in X_{\nu,\rho}$ and $\|Kx\|_Y \leq \delta$. Then there exists a $w \in X$ with $x = |K|^\nu w$ and $\|w\|_X \leq \rho$. The interpolation inequality from Lemma 3.14 for $s = \nu$ and $r = \nu + 1$ together with the properties from Lemma 3.13 then imply that

$$\begin{aligned} \|x\|_X = \||K|^\nu w\|_X &\leq \||K|^{\nu+1} w\|_X^{\frac{\nu}{\nu+1}} \|w\|_X^{\frac{1}{\nu+1}} = \|K|K|^\nu w\|_Y^{\frac{\nu}{\nu+1}} \|w\|_X^{\frac{1}{\nu+1}} \\ &= \|Kx\|_Y^{\frac{\nu}{\nu+1}} \|w\|_X^{\frac{1}{\nu+1}} \leq \delta^{\frac{\nu}{\nu+1}} \rho^{\frac{1}{\nu+1}}. \end{aligned}$$

Taking the supremum over all $x \in X_{\nu,\rho}$ with $\|Kx\|_Y \leq \delta$ yields the claim. □

So far this is only an upper bound, but there always exists at least one sequence for which it is attained.

Theorem 4.12. *Let $K \in \mathcal{K}(X, Y)$ and $\nu, \rho > 0$. Then there exists a null sequence $\{\delta_n\}_{n\in\mathbb{N}}$ with*

$$\varepsilon(X_{\nu,\rho}, \delta_n) = \delta_n^{\frac{\nu}{\nu+1}} \rho^{\frac{1}{\nu+1}}.$$

Proof. Let $\{(\sigma_n, u_n, v_n)\}_{n\in\mathbb{N}}$ be a singular system for K and set $\delta_n := \rho\sigma_n^{\nu+1}$ as well as $x_n := |K|^\nu(\rho v_n)$. Since singular values form a null sequence, we have $\delta_n \to 0$. Furthermore, by construction $x_n \in X_{\nu,\rho}$. It now follows from $\sigma_n = (\rho^{-1}\delta_n)^{\frac{1}{\nu+1}}$ that

$$x_n = \rho|K|^\nu v_n = \rho\sigma_n^\nu v_n = \delta_n^{\frac{\nu}{\nu+1}} \rho^{\frac{1}{\nu+1}} v_n$$

since σ_n^ν is an eigenvalue of $|K|^\nu$ corresponding to the eigenvector v_n. Hence, $\|x_n\|_X = \delta_n^{\frac{\nu}{\nu+1}} \rho^{\frac{1}{\nu+1}}$. Analogously, we obtain that

$$K^*Kx_n = \delta_n^{\frac{\nu}{\nu+1}} \rho^{\frac{1}{\nu+1}} \sigma_n^2 v_n = \delta_n^{\frac{\nu+2}{\nu+1}} \rho^{-\frac{1}{\nu+1}} v_n$$

and thus that

$$\|Kx_n\|_Y^2 = (Kx_n \mid Kx_n)_Y = (K^*Kx_n \mid x_n)_X = \delta_n^2.$$

For all $n \in \mathbb{N}$, we therefore have that

$$\varepsilon(X_{\nu,\rho}, \delta_n) = \sup_{x\in X_{\nu,\rho},\ \|Kx\|_Y\leq\delta_n} \|x\|_X \geq \|x_n\|_X = \delta_n^{\frac{\nu}{\nu+1}} \rho^{\frac{1}{\nu+1}},$$

which together with Theorem 4.11 yields the claimed equality. □

This theorem implies that for a compact operator K with infinite-dimensional range, there can be no regularization method for which the worst-case error can go to zero faster than $\delta_n^{\frac{\nu}{\nu+1}}\rho^{\frac{1}{\nu+1}}$ as $\delta \to 0$ – and even this is only possible under the additional assumption that $x^\dagger \in X_{\nu,\rho}$. In particular, the regularization error always tends to zero more slowly than the data error.

We thus call a regularization method *optimal* (for ν and ρ) if

$$\mathcal{E}(Kx^\dagger,\delta) = \delta^{\frac{\nu}{\nu+1}}\rho^{\frac{1}{\nu+1}} \qquad \text{for all } x^\dagger \in X_{\nu,\rho}$$

and *order optimal* (for ν and ρ) if there exists a constant $c = c(\nu) \geq 1$ such that

$$\mathcal{E}(Kx^\dagger,\delta) \leq c\delta^{\frac{\nu}{\nu+1}}\rho^{\frac{1}{\nu+1}} \qquad \text{for all } x^\dagger \in X_{\nu,\rho}. \tag{4.8}$$

If we allow this constant to depend on $x^\dagger$ – i.e., we are only interested in *convergence rates* – then we set

$$X_\nu := \bigcup_{\rho>0} X_{\nu,\rho} = \mathcal{R}(|K|^\nu)$$

und call a regularization method order optimal for ν if there exists a $c = c(x^\dagger) \geq 1$ such that

$$\mathcal{E}(Kx^\dagger,\delta) \leq c\delta^{\frac{\nu}{\nu+1}} \qquad \text{for all } x^\dagger \in X_\nu.$$

The assumption $x^\dagger \in X_{\nu,\rho}$ is called a *source condition*, and the element $w \in X$ with $|K|^\nu w = x^\dagger$ is sometimes referred to as a *source representer*. Since K is a compact (i.e., smoothing) operator, source conditions are abstract smoothness conditions; e.g., for the integral operator K from Example 3.11, the condition $x \in X_{2,\rho}$ implies that $x = K^*Kw = \int_t^1 \int_0^s w(r)\,dr\,ds$ has second derivative w that is bounded by ρ.

Using the singular value decomposition of K, it is not hard to show that in general the condition $x^\dagger \in X_\nu$ corresponds to a strengthened Picard condition, i.e., that the decay of the Fourier coefficients of y in relation to the singular values of K is faster the larger ν is.

Lemma 4.13. *Let $K \in \mathcal{K}(X,Y)$ have the singular system $\{(\sigma_n,u_n,v_n)\}_{n\in\mathbb{N}}$ and let $y \in \mathcal{R}(K)$. Then $x^\dagger = K^\dagger y \in X_\nu$ if and only if*

$$\sum_{n\in\mathbb{N}} \sigma_n^{-2-2\nu}|\,(y\,|\,u_n)_Y\,|^2 < \infty. \tag{4.9}$$

Proof. From the definition and the representation (3.14), it follows that $K^\dagger y \in X_\nu$ if and only if there exists a $w \in X$ with

$$\sum_{n\in\mathbb{N}} \sigma_n^{-1}\,(y\,|\,u_n)_Y\,v_n = K^\dagger y = |K|^\nu w = \sum_{n\in\mathbb{N}} \sigma_n^\nu\,(w\,|\,v_n)_X\,v_n.$$

Since the v_n form an orthonormal system, we can equate the corresponding coefficients to obtain that

$$\sigma_n^{-1}\,(y\,|\,u_n)_Y = \sigma_n^{\nu}\,(w\,|\,v_n)_X \qquad \text{for all } n \in \mathbb{N}. \tag{4.10}$$

As in the proof of Theorem 3.10, we have that $w \in X$ if and only if $\sum_{n\in\mathbb{N}} |\,(w\,|\,v_n)_X\,|^2$ is finite. Inserting (4.10) now yields (4.9). □

In fact, order optimality already implies the convergence of a regularization method. This is useful since it can be easier to show optimality of a methods than its regularization property (in particular for the discrepancy principle, which motivates the slightly complicated statement of the following theorem).

Theorem 4.14. *Let $K \in \mathcal{K}(X, Y)$ with $\mathcal{R}(K)$ dense in Y, $\{R_\alpha\}_{\alpha>0}$ be a regularization, and $\alpha(\delta, y^\delta)$ be a parameter choice rule. If there exists a $\tau_0 \geq 1$ such that R_α together with $\alpha_\tau := \alpha(\tau\delta, y^\delta)$ for all $\tau > \tau_0$ satisfies the condition (4.8) for some $\nu > 0$ and all $\rho > 0$, then (R_α, α_τ) is a regularization method for all $\tau > \tau_0$.*

Proof. We have to show that the *uniform* convergence of the worst-case error for all $x^\dagger \in X_{\nu,\rho}$ implies the *pointwise* convergence for all $x^\dagger \in \mathcal{R}(K^\dagger)$. For this, we construct a suitable $x_N \in X_{\nu,\rho}$, insert it into the error estimate, and apply the order optimality.

Let therefore $y \in \mathcal{D}(K^\dagger) = \mathcal{R}(K)$ and $x^\dagger = K^\dagger y$ (and hence $Kx^\dagger = y$). Furthermore, let $\{(\sigma_n, u_n, v_n)\}_{n\in\mathbb{N}}$ be a singular system of K. We now define for $N \in \mathbb{N}$

$$x_N := \sum_{n=1}^{N} \left(x^\dagger \,\middle|\, v_n\right)_X v_n$$

and

$$\begin{aligned} y_N := Kx_N &= \sum_{n=1}^{N} \left(x^\dagger \,\middle|\, v_n\right)_X Kv_n = \sum_{n=1}^{N} \left(x^\dagger \,\middle|\, v_n\right)_X \sigma_n u_n \\ &= \sum_{n=1}^{N} \left(x^\dagger \,\middle|\, K^* u_n\right)_X u_n = \sum_{n=1}^{N} (y\,|\,u_n)_X\, u_n. \end{aligned}$$

Since $\{u_n\}_{n\in\mathbb{N}}$ is an orthonormal basis of $\overline{\mathcal{R}(K)}$ and $\{v_n\}_{n\in\mathbb{N}}$ is an orthonormal basis of $\overline{\mathcal{R}(K^*)} = \mathcal{N}(K)^\perp$, we can represent $x^\dagger = K^\dagger y \in \mathcal{N}(K)^\perp$ and $y = Kx^\dagger \in \mathcal{R}(K)$ as

$$x^\dagger = \sum_{n\in\mathbb{N}} \left(x^\dagger \,\middle|\, v_n\right)_X v_n, \qquad y = \sum_{n\in\mathbb{N}} (y\,|\,u_n)_Y\, u_n.$$

From this we obtain

$$\|x^\dagger - x_N\|_X^2 = \sum_{n=N+1}^{\infty} \left|\left(x^\dagger \,\middle|\, v_n\right)_X\right|^2$$

and

$$\|y - y_N\|_Y^2 = \sum_{n=N+1}^{\infty} \left|(y \mid u_n)_Y\right|^2 = \sum_{n=N+1}^{\infty} \sigma_n^2 \left|\left(x^\dagger \mid v_n\right)_X\right|^2 \tag{4.11}$$
$$\leq \sigma_N^2 \sum_{n=N+1}^{\infty} \left|\left(x^\dagger \mid v_n\right)_X\right|^2 = \sigma_N^2 \|x^\dagger - x_N\|_X^2$$

since $\{\sigma_n\}_{n\in\mathbb{N}}$ is a decreasing null sequence. In particular, $x_N \to x^\dagger$ and $y_N \to y$ as $N \to \infty$.

By construction, $y_N \in \mathcal{R}(K)$ and $x_N \in \mathcal{N}(K)^\perp$ and therefore $x_N = K^\dagger y_N$. From Lemma 4.13 we thus deduce that $x_N \in X_\nu$ for all $\nu > 0$, since it follows from $(y_N \mid u_n)_Y = 0$ for $n > N$ that the series in (4.9) is finite. Hence there exists an $w_N \in X$ with $x_N = |K|^\nu w_N$, i.e.,

$$\sum_{n=1}^{N} \left(x^\dagger \mid v_n\right)_X v_n = x_N = |K|^\nu w_N = \sum_{n\in\mathbb{N}} \sigma_n^\nu \left(w_N \mid v_n\right)_X v_n.$$

As $\mathcal{R}(K)$ is dense in Y, the range of K can not be infinite-dimensional, which implies that $\sigma_n > 0$ for all $n \in \mathbb{N}$. Since the v_n form an orthonormal system, we thus obtain that

$$(w_N \mid v_n)_X = \begin{cases} \sigma_n^{-\nu} \left(x^\dagger \mid v_n\right)_X & n \leq N, \\ 0 & n > N, \end{cases}$$

and hence that

$$\|w_N\|_X^2 = \sum_{n=1}^{N} |(w_N \mid v_n)_X|^2 = \sum_{n=1}^{N} \sigma_n^{-2\nu} \left|\left(x^\dagger \mid v_n\right)_X\right|^2$$
$$\leq \sigma_N^{-2\nu} \sum_{n\in\mathbb{N}} \left|\left(x^\dagger \mid v_n\right)_X\right|^2 = \sigma_N^{-2\nu} \|x^\dagger\|_X^2.$$

This implies that $x_N \in X_{\nu,\rho}$ with $\rho = \sigma_N^{-\nu} \|x^\dagger\|_X$.

Let now $y^\delta \in B_\delta(y)$ and $\tau > \tau_0 \geq 1$ and choose $N(\delta)$ such that

$$\sigma_{N(\delta)} \|x^\dagger - x_{N(\delta)}\|_X \leq \frac{\tau - \tau_0}{\tau + \tau_0} \delta < \sigma_{N(\delta)-1} \|x^\dagger - x_{N(\delta)-1}\|_X \tag{4.12}$$

(which is possible since both $\{\sigma_N\}_{N\in\mathbb{N}}$ and $\{\|x_N - x^\dagger\|_X\}_{N\in\mathbb{N}}$ are decreasing null sequences). We then obtain from (4.11) with $N = N(\delta)$ that

$$\|y^\delta - y_N\|_Y \leq \|y^\delta - y\|_Y + \|y - y_N\|_Y \leq \delta + \sigma_N \|x^\dagger - x_N\|_X$$
$$\leq \left(1 + \frac{\tau - \tau_0}{\tau + \tau_0}\right) \delta =: \tilde{\delta}.$$

Hence if y^δ is a noisy measurement for the exact data y with noise level δ, then y^δ is also a noisy measurement for y_N with noise level $\tilde{\delta}$. Setting $\tilde{\tau} := \frac{1}{2}(\tau + \tau_0) > \tau_0$, we thus have $\tilde{\tau}\tilde{\delta} = \tau\delta$ and therefore

$$\alpha_{\tilde{\tau}}(\tilde{\delta}, y^\delta) = \alpha(\tilde{\tau}\tilde{\delta}, y^\delta) = \alpha(\tau\delta, y^\delta) = \alpha_\tau(\delta, y^\delta),$$

i.e., the parameter choice rules α_τ for y and $\alpha_{\tilde\tau}$ for y_N coincide for given y^δ. The order optimality (4.8) of $(R_\alpha, \alpha_{\tilde\tau})$ for $x_N \in X_{\nu,\rho}$ thus implies that

$$\|R_{\alpha_\tau(\delta,y^\delta)}y^\delta - x_N\|_X = \|R_{\alpha_{\tilde\tau}(\tilde\delta,y^\delta)}y^\delta - K^\dagger y_N\|_X \leq \mathcal{E}(y_N,\tilde\delta) \leq c\tilde\delta^{\frac{\nu}{\nu+1}}\left(\sigma_N^{-\nu}\|x^\dagger\|_X\right)^{\frac{1}{\nu+1}}$$
$$=: c_{\tau,\nu}\left(\frac{\delta}{\sigma_N}\right)^{\frac{\nu}{\nu+1}}\|x^\dagger\|_X^{\frac{1}{\nu+1}}.$$

We thus have that

$$\begin{aligned}\|R_{\alpha_\tau(\delta,y^\delta)}y^\delta - x^\dagger\|_X &\leq \|R_{\alpha_\tau(\delta,y^\delta)}y^\delta - x_{N(\delta)}\|_X + \|x_{N(\delta)} - x^\dagger\|_X\\ &\leq c_{\tau,\nu}\left(\frac{\delta}{\sigma_{N(\delta)}}\right)^{\frac{\nu}{\nu+1}}\left(\|x^\dagger\|_X\right)^{\frac{1}{\nu+1}} + \|x_{N(\delta)} - x^\dagger\|_X,\end{aligned}$$

and it remains to show that both $\delta\sigma_{N(\delta)}^{-1} \to 0$ and $x_{N(\delta)} \to x^\dagger$ as $\delta \to 0$. Since $N(\delta)$ is an increasing function of δ, we only have to distinguish two cases:

(i) $N(\delta)$ is bounded and therefore convergent, i.e., there exists an $N_0 < \infty$ with $N(\delta) \to N_0$ as $\delta \to 0$. In this case, we obviously have that $\delta\sigma_{N(\delta)}^{-1} \leq \delta\sigma_{N_0}^{-1} \to 0$. Furthermore, the choice of $N(\delta)$ according to (4.12) implies that

$$\sigma_{N_0}\|x^\dagger - x_{N_0}\|_X = \lim_{\delta\to 0}\sigma_{N(\delta)}\|x^\dagger - x_{N(\delta)}\|_X \leq \lim_{\delta\to 0}\frac{\tau-\tau_0}{\tau+\tau_0}\delta = 0$$

and hence that $x_{N(\delta)} \to x_{N_0} = x^\dagger$ due to $\sigma_{N_0} > 0$.

(ii) $N(\delta)$ is unbounded, i.e., $N(\delta) \to \infty$ as $\delta \to 0$. This immediately implies that $x_{N(\delta)} \to x^\dagger$. Furthermore, we obtain from (4.12) that

$$\frac{\delta}{\sigma_{N(\delta)}} < \frac{\tau+\tau_0}{\tau-\tau_0}\frac{\sigma_{N(\delta)-1}}{\sigma_{N(\delta)}}\|x_{N(\delta)-1} - x^\dagger\|_X \to 0$$

since the second factor on the right-hand side stays bounded due to $\sigma_{N(\delta)} \to 0$.

Hence, $R_{\alpha_\tau(\delta,y^\delta)}y^\delta \to x^\dagger$ for all $y \in \mathcal{D}(K^\dagger)$ and $y^\delta \in B_\delta(y)$, and thus (R_α, α_τ) is a regularization method. ◻

Finally, we remark that it is possible to formulate weaker source conditions using more general *index functions* ψ than powers. One example are *logarithmic source conditions* of the form $x^\dagger \in \mathcal{R}(-\ln|K|)$ that are appropriate for exponentially ill-posed problems; see, e.g., [Hohage 2000]. In fact, it is possible to show that for every $x^\dagger \in X$ there exists an index function ψ with $x^\dagger \in \mathcal{R}(\psi(|K|))$ for which the worst-case error can be bounded in terms of ψ; see [Mathé & Hofmann 2008].

5 SPECTRAL REGULARIZATION

As we have seen, regularizing an ill-posed operator equation $Tx = y$ consists in replacing the (unbounded) Moore–Penrose inverse $T^\dagger$ by a family $\{R_\alpha\}_{\alpha>0}$ of operators that for $\alpha > 0$ are continuous on Y and for $\alpha \to 0$ converge pointwise on $\mathcal{D}(T^\dagger)$ to $T^\dagger$. For a compact operator $K \in \mathcal{K}(X, Y)$, such regularizations can be constructed using the singular value decomposition together with the fact that by Corollary 3.6 we have for $y \in \mathcal{D}(K^\dagger)$ that

$$K^\dagger y = (K^*K)^\dagger K^* y.$$

Let therefore $\{(\sigma_n, u_n, v_n)\}_{n\in\mathbb{N}}$ be a singular system of K. By construction, $\{(\sigma_n^2, v_n, v_n)\}_{n\in\mathbb{N}}$ is then a singular system of K^*K, and Theorem 3.10 yields that

$$\begin{aligned}(K^*K)^\dagger K^* y &= \sum_{n\in\mathbb{N}} \sigma_n^{-2} \,(K^* y \,|\, v_n)_X \, v_n = \sum_{n\in\mathbb{N}} \sigma_n^{-2}\sigma_n \,(y \,|\, u_n)_Y \, v_n \\ &= \sum_{n\in\mathbb{N}} \varphi(\sigma_n^2)\sigma_n \,(y \,|\, u_n)_Y \, v_n\end{aligned}$$

for $\varphi(\lambda) = \lambda^{-1}$. The unboundedness of $K^\dagger$ is thus due to the fact that φ is unbounded on $(0, \|K^*K\|_{\mathbb{L}(X,X)}]$ and that $\{\sigma_n\}_{n\in\mathbb{N}}$ is a null sequence. To obtain a regularization, we therefore replace φ by a family $\{\varphi_\alpha\}_{\alpha>0}$ of *bounded* functions that converge pointwise to φ. Here and throughout the following, we set $\kappa := \|K\|^2_{\mathbb{L}(X,Y)} = \|K^*K\|_{\mathbb{L}(X,X)}$ for brevity.

Definition 5.1. Let $\{\varphi_\alpha\}_{\alpha>0}$ be a family of piecewise continuous and bounded functions $\varphi_\alpha : [0, \kappa] \to \mathbb{R}$. If

(i) $\lim_{\alpha\to 0} \varphi_\alpha(\lambda) = \frac{1}{\lambda}$ for all $\lambda \in (0, \kappa]$ and

(ii) $\lambda|\varphi_\alpha(\lambda)| \leq C_\varphi$ for some $C_\varphi > 0$ and all $\lambda \in (0, \kappa]$ and $\alpha > 0$,

then $\{\varphi_\alpha\}_{\alpha>0}$ is called a *(regularizing) filter*.

The idea is now to take $R_\alpha := \varphi_\alpha(K^*K)K^*$ as a regularization operator, i.e., to set for $y \in Y$

$$\begin{aligned}R_\alpha y = \varphi_\alpha(K^*K)K^* y &= \sum_{n\in\mathbb{N}} \varphi_\alpha(\sigma_n^2) \,(K^* y \,|\, v_n)_Y \, v_n + \varphi_\alpha(0) P_{\mathcal{N}} K^* y \\ &= \sum_{n\in\mathbb{N}} \varphi_\alpha(\sigma_n^2)\sigma_n \,(y \,|\, u_n)_Y \, v_n\end{aligned}$$

since $K^* y \in \overline{\mathcal{R}(K^*)} = \mathcal{N}(K)^\perp$. This approach covers several prototypical regularizations.

Example 5.2. (i) The *truncated singular value decomposition* corresponds to the choice

$$\varphi_\alpha(\lambda) = \begin{cases} \frac{1}{\lambda} & \text{if } \lambda \geq \alpha, \\ 0 & \text{else.} \end{cases} \tag{5.1}$$

Obviously, φ_α is bounded (by $\frac{1}{\alpha}$) and piecewise continuous, converges for $\lambda > 0$ to $\frac{1}{\lambda}$ as $\alpha \to 0$, and satisfies the boundedness condition for $C_\varphi = 1$. The corresponding regularization operator is given by

$$R_\alpha y = \sum_{n\in\mathbb{N}} \varphi_\alpha(\sigma_n^2)\sigma_n \, (y \,|\, u_n)_Y \, v_n = \sum_{\sigma_n \geq \sqrt{\alpha}} \frac{1}{\sigma_n} \, (y \,|\, u_n)_Y \, v_n, \tag{5.2}$$

which also explains the name. We will revisit this example throughout this chapter.

(ii) The *Tikhonov regularization* corresponds to the choice

$$\varphi_\alpha(\lambda) = \frac{1}{\lambda + \alpha}.$$

Again, φ_α is bounded (by $\frac{1}{\alpha}$) and continuous, converges for $\lambda > 0$ to $\frac{1}{\lambda}$ as $\alpha \to 0$, and satisfies the boundedness condition for $C_\varphi = 1$. The corresponding regularization operator is given by

$$R_\alpha y = \sum_{n\in\mathbb{N}} \frac{\sigma_n}{\sigma_n^2 + \alpha} \, (y \,|\, u_n)_Y \, v_n.$$

However, the regularization $\varphi_\alpha(K^*K)K^*y$ can be computed without the aid of a singular value decomposition; we will treat this in detail in Chapter 6.

(iii) The *Landweber regularization* corresponds to the choice

$$\varphi_\alpha(\lambda) = \frac{1 - (1 - \omega\lambda)^{1/\alpha}}{\lambda}$$

for a suitable $\omega > 0$. If ω is small enough, one can show that this choice satisfies the definition of a regularizing filter. But here as well we can give a (more intuitive) characterization of the corresponding regularization operator without singular value decompositions; we therefore postpone its discussion to Chapter 7.

5.1 REGULARIZATION

We first show that if $\{\varphi_\alpha\}_{\alpha>0}$ is a regularizing filter, then $R_\alpha := \varphi_\alpha(K^*K)K^*$ defines indeed a regularization $\{R_\alpha\}_{\alpha>0}$ of $K^\dagger$. For this we will need the following three fundamental lemmas, which will be used throughout this chapter.

Lemma 5.3. *Let $\{\varphi_\alpha\}_{\alpha>0}$ be a regularizing filter. Then*

$$\|KR_\alpha\|_{\mathbb{L}(Y,Y)} \leq \sup_{n\in\mathbb{N}} |\varphi_\alpha(\sigma_n^2)|\sigma_n^2 \leq C_\varphi \qquad \textit{for all } \alpha > 0.$$

Proof. For all $y \in Y$ and $\alpha > 0$, we have that (compare (3.11))

$$\begin{aligned} KR_\alpha y = K\varphi_\alpha(K^*K)K^*y &= \sum_{n\in\mathbb{N}} \varphi_\alpha(\sigma_n^2)\sigma_n \left(y \mid u_n\right)_y Kv_n \\ &= \sum_{n\in\mathbb{N}} \varphi_\alpha(\sigma_n^2)\sigma_n^2 \left(y \mid u_n\right)_y u_n. \end{aligned} \tag{5.3}$$

Together with the Bessel inequality (2.2), this implies that

$$\begin{aligned} \|KR_\alpha y\|_Y^2 = \sum_{n\in\mathbb{N}} |\varphi_\alpha(\sigma_n^2)\sigma_n^2 \left(y \mid u_n\right)_y|^2 &\leq \sup_{n\in\mathbb{N}} |\varphi_\alpha(\sigma_n^2)\sigma_n^2|^2 \sum_{n\in\mathbb{N}} |\left(y \mid u_n\right)_Y|^2 \\ &\leq \sup_{n\in\mathbb{N}} |\varphi_\alpha(\sigma_n^2)\sigma_n^2|^2 \|y\|_Y^2. \end{aligned}$$

The second inequality now follows from the fact that $0 < \sigma_n^2 \leq \sigma_1^2 = \|K^*K\|_{\mathbb{L}(X,X)} = \kappa$ together with the boundedness condition (ii) of regularizing filters. ◻

Lemma 5.4. *Let $\{\varphi_\alpha\}_{\alpha>0}$ be a regularizing filter. Then*

$$\|R_\alpha\|_{\mathbb{L}(Y,X)} \leq \sqrt{C_\varphi} \sup_{\lambda\in(0,\kappa]} \sqrt{|\varphi_\alpha(\lambda)|} \qquad \textit{for all } \alpha > 0.$$

In particular, $R_\alpha : Y \to X$ is continuous for all $\alpha > 0$.

Proof. For all $y \in Y$ and $\alpha > 0$, it follows from Lemma 5.3 and $\sigma_n v_n = K^* u_n$ that

$$\begin{aligned} \|R_\alpha y\|_X^2 = \left(R_\alpha y \mid R_\alpha y\right)_X &= \sum_{n\in\mathbb{N}} \varphi_\alpha(\sigma_n^2)\sigma_n \left(y \mid u_n\right)_Y \left(R_\alpha y \mid v_n\right)_X \\ &= \sum_{n\in\mathbb{N}} \varphi_\alpha(\sigma_n^2) \left(y \mid u_n\right)_Y \left(KR_\alpha y \mid u_n\right)_Y \\ &\leq \sup_{n\in\mathbb{N}} |\varphi_\alpha(\sigma_n^2)| \left(KR_\alpha y \,\middle|\, \textstyle\sum_{n\in\mathbb{N}} \left(y \mid u_n\right)_Y u_n\right)_Y \\ &\leq \sup_{n\in\mathbb{N}} |\varphi_\alpha(\sigma_n^2)| \, \|KR_\alpha y\|_X \|P_{\overline{\mathcal{R}(K^*)}} y\|_Y \\ &\leq \sup_{n\in\mathbb{N}} |\varphi_\alpha(\sigma_n^2)| \, C_\varphi \|y\|_Y^2. \end{aligned}$$

Taking the supremum over all $y \in Y$ and using the boundedness of φ_α now yields the claim. ◻

Finally, the third "fundamental lemma of spectral regularization" gives a spectral representation of the approximation error.

Lemma 5.5. *Let $\{\varphi_\alpha\}_{\alpha>0}$ be a regularizing filter and $y \in \mathcal{D}(K^\dagger)$. Then for every $x^\dagger := K^\dagger y$,*

$$K^\dagger y - R_\alpha y = \sum_{n\in\mathbb{N}} r_\alpha(\sigma_n^2) \left(x^\dagger \,\middle|\, v_n\right)_X v_n,$$

where $r_\alpha(\lambda) := 1 - \lambda\varphi_\alpha(\lambda)$ satisfies

$$\begin{aligned} \lim_{\alpha\to 0} r_\alpha(\lambda) &= 0 && \text{for all } \lambda \in (0,\kappa], \\ |r_\alpha(\lambda)| &\le 1 + C_\varphi && \text{for all } \lambda \in (0,\kappa] \text{ and } \alpha > 0. \end{aligned}$$

Proof. Since $K^*Kx^\dagger = K^*y$ by Corollary 3.6, we can write

$$R_\alpha y = \varphi_\alpha(K^*K)K^*y = \varphi_\alpha(K^*K)K^*Kx^\dagger,$$

and the definition of r_α immediately yields that

$$K^\dagger y - R_\alpha y = (\mathrm{Id} - \varphi_\alpha(K^*K)K^*K)x^\dagger = r_\alpha(K^*K)x^\dagger = \sum_{n\in\mathbb{N}} r_\alpha(\sigma_n^2) \left(x^\dagger \,\middle|\, v_n\right)_X v_n.$$

The remaining claims follow from the corresponding properties of regularizing filters. □

We now have everything at hand to show the pointwise convergence and thus the regularization property of $\{R_\alpha\}_{\alpha>0}$.

Theorem 5.6. *Let $\{\varphi_\alpha\}_{\alpha>0}$ be a regularizing filter. Then*

$$\lim_{\alpha\to 0} R_\alpha y = K^\dagger y \qquad \text{for all } y \in \mathcal{D}(K^\dagger),$$

i.e., $\{R_\alpha\}_{\alpha>0}$ is a regularization.

Furthermore, if $K^\dagger$ is not continuous, then $\lim_{\alpha\to 0} \|R_\alpha y\|_X = \infty$ for all $y \notin \mathcal{D}(K^\dagger)$.

Proof. Let $y \in \mathcal{D}(K^\dagger)$ and $x^\dagger = K^\dagger y$. Lemma 5.5 then yields that

$$\|K^\dagger y - R_\alpha y\|_X^2 = \sum_{n\in\mathbb{N}} r_\alpha(\sigma_n^2)^2 \left|\left(x^\dagger \,\middle|\, v_n\right)_X\right|^2.$$

To show that the right-hand side tends to zero as $\alpha \to 0$, we split the series into a finite sum, for which we can use the convergence of r_α and the boundedness of the Fourier coefficients, and a remainder term, for which we argue vice versa.

Let therefore $\varepsilon > 0$ be arbitrary. Then we first obtain from the Bessel inequality an $N \in \mathbb{N}$ with

$$\sum_{n=N+1}^{\infty} \left|\left(x^\dagger \,\middle|\, v_n\right)_X\right|^2 < \frac{\varepsilon^2}{2(1+C_\varphi)^2}.$$

Furthermore, the pointwise convergence of $\{r_\alpha\}_{\alpha>0}$ – which is uniform on the finite set $\{\sigma_1^2, \dots, \sigma_N^2\}$ – yields an $\alpha_0 > 0$ with

$$r_\alpha(\sigma_n^2)^2 < \frac{\varepsilon^2}{2\|x^\dagger\|_X^2} \qquad \text{for all } n \leq N \text{ and } \alpha < \alpha_0.$$

We thus have for all $\alpha < \alpha_0$ that

$$\begin{aligned}
\|K^\dagger y - R_\alpha y\|_X^2 &= \sum_{n=1}^{N} r_\alpha(\sigma_n^2)^2 \left|\left(x^\dagger \,\middle|\, v_n\right)_X\right|^2 + \sum_{n=N+1}^{\infty} r_\alpha(\sigma_n^2)^2 \left|\left(x^\dagger \,\middle|\, v_n\right)_X\right|^2 \\
&\leq \frac{\varepsilon^2}{2\|x^\dagger\|_X^2} \sum_{n=1}^{N} \left|\left(x^\dagger \,\middle|\, v_n\right)_X\right|^2 + (1+C_\varphi)^2 \frac{\varepsilon^2}{2(1+C_\varphi)^2} \\
&\leq \frac{\varepsilon^2}{2} + \frac{\varepsilon^2}{2} = \varepsilon^2,
\end{aligned}$$

i.e., $\|K^\dagger y - R_\alpha y\|_X \to 0$ as $\alpha \to 0$. Together with the continuity of R_α for $\alpha > 0$ from Lemma 5.4, this implies by Definition 4.1 that $\{R_\alpha\}_{\alpha>0}$ is a regularization.

Finally, the divergence for $y \notin \mathcal{D}(K^\dagger)$ follows from Theorem 4.3 and Lemma 5.3. □

In particular, the truncated singular value decomposition, the Tikhonov regularization, and the Landweber regularization from Example 5.2 define regularizations.

5.2 PARAMETER CHOICE AND CONVERGENCE RATES

We now investigate which parameter choice rules α will for a given filter φ_α lead to a convergent (and order optimal) regularization method (R_α, α). To keep the notation concise, we will in the following write $x^\dagger := K^\dagger y$, $x_\alpha := R_\alpha y$ for $y \in \mathcal{D}(K^\dagger)$, and $x_\alpha^\delta := R_\alpha y^\delta$ for $y^\delta \in B_\delta(y)$.

A PRIORI CHOICE RULES

By Theorem 4.6, every a priori choice rule that satisfies $\alpha(\delta) \to 0$ and $\delta\|R_\alpha\|_{\mathbb{L}(Y,X)} \to 0$ and $\delta \to 0$ leads to a regularization method (R_α, α). Together with Lemma 5.4, this leads to a condition on φ_α and thus on α.

Example 5.7 (truncated singular value decomposition). Let $K \in \mathcal{K}(X,Y)$ have the singular system $\{(\sigma_n, u_n, v_n)\}_{n\in\mathbb{N}}$. Then we have for φ_α as in (5.1) that

$$\|R_\alpha\|_{\mathbb{L}(Y,X)} \le \sqrt{C_\varphi} \sup_{n\in\mathbb{N}} \sqrt{|\varphi_\alpha(\sigma_n^2)|} = \frac{1}{\sqrt{\alpha}}.$$

This yields a condition on the minimal singular value that we can include in (5.2) for given $\delta > 0$: Choosing $n(\delta)$ with

$$n(\delta) \to \infty, \qquad \frac{\delta}{\sigma_{n(\delta)}} \to 0 \qquad \text{as } \delta \to 0,$$

the truncated singular value decomposition together with the parameter choice rule $\alpha(\delta) := \sigma_{n(\delta)}^2$ becomes a regularization method.

In particular, this holds for the choice $\alpha(\delta) := \sigma_{n(\delta)}^2 \ge \delta > \sigma_{n(\delta)+1}^2$, which satisfies

$$x_{\alpha(\delta)}^\delta = \sum_{\sigma_n \ge \sqrt{\delta}} \frac{1}{\sigma_n} \left(y^\delta \,\middle|\, u_n\right)_Y v_n \to \sum_{n\in\mathbb{N}} \frac{1}{\sigma_n} (y \mid u_n)_Y v_n = x^\dagger \qquad \text{as } \delta \to 0.$$

We now consider convergence rates under the source condition $x^\dagger \in X_{\nu,\rho}$ for $\nu, \rho > 0$. For this, we proceed as in the proof of Theorem 5.6 and first show that

$$\omega_\nu(\alpha) := \sup_{\lambda\in(0,\kappa]} \lambda^{\nu/2} |r_\alpha(\lambda)|$$

is an upper bound for the approximation error.

Lemma 5.8. *Let $y \in \mathcal{D}(K^\dagger)$ and $x^\dagger \in X_{\nu,\rho}$ for some $\nu, \rho > 0$. Then we have for all $\alpha > 0$ that*

$$\|x_\alpha - x^\dagger\|_X \le \omega_\nu(\alpha)\rho, \tag{5.4}$$

$$\|Kx_\alpha - Kx^\dagger\|_Y \le \omega_{\nu+1}(\alpha)\rho. \tag{5.5}$$

Proof. By definition, for $x^\dagger \in X_{\nu,\rho}$ there exists a $w \in X$ with $x^\dagger = |K|^\nu w = (K^*K)^{\nu/2} w$ and $\|w\|_X \le \rho$. It then follows from Lemma 5.5 that

$$\begin{aligned} x^\dagger - x_\alpha &= r_\alpha(K^*K)x^\dagger = r_\alpha(K^*K)(K^*K)^{\nu/2} w \\ &= \sum_{n\in\mathbb{N}} r_\alpha(\sigma_n^2)\sigma_n^\nu (w \mid v_n)_X v_n \end{aligned}$$

and hence that

$$\begin{aligned} \|x_\alpha - x^\dagger\|_X^2 &= \sum_{n\in\mathbb{N}} |r_\alpha(\sigma_n^2)|^2 \sigma_n^{2\nu} |(w \mid v_n)_X|^2 \\ &\le \omega_\nu(\alpha)^2 \sum_{n\in\mathbb{N}} |(w \mid v_n)_X|^2 \le \omega_\nu(\alpha)^2 \|w\|_X^2 \le \omega_\nu(\alpha)^2\rho^2. \end{aligned}$$

Furthermore, Lemma 3.13 (iii) yields

$$\|Kx_\alpha - Kx^\dagger\|_Y = \|K(x_\alpha - x^\dagger)\|_Y = \||K|(x_\alpha - x^\dagger)\|_X.$$

From this together with

$$\begin{aligned}|K|(x^\dagger - x_\alpha) &= (K^*K)^{1/2} r_\alpha(K^*K)(K^*K)^{\nu/2} w\\ &= \sum_{n\in\mathbb{N}} \sigma_n r_\alpha(\sigma_n^2)\sigma_n^\nu \,(w\,|\,v_n)_X\, v_n\end{aligned}$$

and $|r_\alpha(\sigma_n^2)\sigma_n^{\nu+1}|^2 \leq \omega_{\nu+1}(\alpha)^2$, we similarly obtain the second estimate. □

We now have everything at hand to show convergence rates.

Theorem 5.9. *Let $y \in \mathcal{D}(K^\dagger)$ and $x^\dagger = K^\dagger y \in X_{\nu,\rho}$ for some $\nu, \rho > 0$. If $\alpha(\delta)$ is an a priori choice rule with*

$$c\left(\frac{\delta}{\rho}\right)^{\frac{2}{\nu+1}} \leq \alpha(\delta) \leq C\left(\frac{\delta}{\rho}\right)^{\frac{2}{\nu+1}} \qquad \text{for } C > c > 0 \tag{5.6}$$

and the filter $\{\varphi_\alpha\}_{\alpha>0}$ satisfies for some $C_\nu > 0$ the conditions

$$\sup_{\lambda\in(0,\kappa]} |\varphi_\alpha(\lambda)| \leq C_\varphi \alpha^{-1}, \tag{5.7}$$

$$\omega_\nu(\alpha) \leq C_\nu \alpha^{\nu/2}, \tag{5.8}$$

then (R_α, α) is a (for this ν and all ρ) order optimal regularization method.

Proof. By Theorem 4.14, it suffices to show order optimality. We again use the decomposition (4.2) into data error and approximation error: For given $\delta > 0$ and $y^\delta \in B_\delta(y)$,

$$\|x^\delta_{\alpha(\delta)} - x^\dagger\|_X \leq \delta \|R_{\alpha(\delta)}\|_{\mathbb{L}(Y,X)} + \|x_{\alpha(\delta)} - x^\dagger\|_X.$$

By Lemma 5.4 and the assumption (5.7), we have that

$$\|R_{\alpha(\delta)}\|_{\mathbb{L}(Y,X)} \leq \sqrt{C_\varphi}\sqrt{C_\varphi \alpha(\delta)^{-1}} \leq C_\varphi \alpha(\delta)^{-1/2}.$$

Similarly, it follows from Lemma 5.8 and the assumption (5.8) that

$$\|x_{\alpha(\delta)} - x^\dagger\|_X \leq \omega_\nu(\alpha(\delta))\rho \leq C_\nu \alpha(\delta)^{\nu/2}\rho.$$

Together with the parameter choice rule (5.6), we obtain

$$\begin{aligned}\|x^\delta_{\alpha(\delta)} - x^\dagger\|_X &\leq C_\varphi \alpha(\delta)^{-1/2}\delta + C_\nu \alpha(\delta)^{\nu/2}\rho\\ &\leq C_\varphi c^{-1/2}\delta^{-\frac{1}{\nu+1}}\rho^{\frac{1}{\nu+1}}\delta + C_\nu C^{\nu/2}\delta^{\frac{\nu}{\nu+1}}\rho^{-\frac{\nu}{\nu+1}}\rho\\ &= (C_\varphi c^{-1/2} + C_\nu C^{\nu/2})\delta^{\frac{\nu}{\nu+1}}\rho^{\frac{1}{\nu+1}}\end{aligned} \tag{5.9}$$

and thus the order optimality. □

Hence, to show for a given filter φ_α the order optimality for some $\nu > 0$, it suffices to verify for this ν the condition (5.8) (as well as for φ_α the condition (5.7)). The maximal $\nu_0 > 0$, for which all $\nu \in (0, \nu_0]$ satisfy the condition (5.8), is called the *qualification* of the filter.

Example 5.10 (truncated singular value decomposition). It follows from (5.1) that

$$\sup_{\lambda\in(0,\kappa]} |\varphi_\alpha(\lambda)| \le \alpha^{-1},$$

and hence this filter satisfies (5.7) with $C_\varphi = 1$.

Furthermore, for all $\nu > 0$ and $\lambda \in (0, \kappa]$,

$$\lambda^{\nu/2}|r_\alpha(\lambda)| = \lambda^{\nu/2}|1 - \lambda\varphi_\alpha(\lambda)| = \begin{cases} 0 & \text{if } \lambda \ge \alpha, \\ \lambda^{\nu/2} & \text{if } \lambda < \alpha. \end{cases}$$

Hence for all $\alpha \in (0, \kappa]$,

$$\omega_\nu(\alpha) = \sup_{\lambda\in(0,\kappa]} \lambda^{\nu/2}|r_\alpha(\lambda)| \le \max\{0, \alpha^{\nu/2}\} = \alpha^{\nu/2},$$

and the condition (5.8) is therefore satisfied for all $\nu > 0$ with $C_\nu = 1$. This shows that the truncated singular value decomposition is order optimal for all $\nu > 0$ and thus has *infinite qualification*.

A POSTERIORI CHOICE RULES

We again consider the discrepancy principle: Fix $\tau > 1$ and choose $\alpha(\delta, y^\delta)$ such that

$$\|Kx^\delta_{\alpha(\delta,y^\delta)} - y^\delta\|_Y \le \tau\delta < \|Kx^\delta_\alpha - y^\delta\|_Y \qquad \text{for all } \alpha > \alpha(\delta, y^\delta). \tag{5.10}$$

As before, we assume that $\mathcal{R}(K)$ is dense in Y. Under additional assumptions on the continuity of the mapping $\alpha \mapsto \|Kx^\delta_\alpha - y^\delta\|_Y$, it is then possible to show that such an $\alpha(\delta, y^\delta)$ always exists (compare Theorem 4.7 with Lemma 5.3). To show that the discrepancy principle leads to an order optimal regularization method, we again apply Theorem 4.14, for which we have to take the discrepancy principle as a parameter choice rule $\alpha_\tau = \alpha(\tau\delta, y^\delta)$.

Theorem 5.11. *Let $\{\varphi_\alpha\}_{\alpha>0}$ be a filter with qualification $\nu_0 > 0$ (i.e., satisfying (5.7) and (5.8) for all $\nu \in (0, \nu_0]$), and let*

$$\tau > \sup_{\alpha>0,\,\lambda\in(0,\kappa]} |r_\alpha(\lambda)| =: C_r. \tag{5.11}$$

Then the discrepancy principle defines for all $\nu \in (0, \nu_0 - 1]$ an order optimal regularization method (R_α, α_τ).

Proof. We first observe that due to $|r_\alpha(\lambda)| \le 1 + C_\varphi$ for all $\alpha > 0$ and $\lambda \in (0, \kappa]$, there always exists a $\tau > 1$ satisfying (5.11).

Let now $y \in \mathcal{R}(K)$, $x^\dagger = K^\dagger y \in X_{\nu,\rho}$ for some $\nu \in (0, \nu_0 - 1]$ and $\rho > 0$, and $y^\delta \in B_\delta(y)$. We again use for $x_\alpha^\delta := x^\delta_{\alpha(\delta,y^\delta)}$ and $x_\alpha := x_{\alpha(\delta,y^\delta)}$ the decomposition

$$\|x_\alpha^\delta - x^\dagger\|_X \le \|x_\alpha - x^\dagger\|_X + \|x_\alpha - x_\alpha^\delta\|_X \tag{5.12}$$

and estimate the terms on the right-and side separately.

For the first term, we again use the representation of the approximation errors from Lemma 5.5 as well as the source condition $x^\dagger = |K|^\nu w$ to obtain

$$\begin{aligned} x^\dagger - x_\alpha &= \sum_{n\in\mathbb{N}} r_\alpha(\sigma_n^2)\sigma_n^\nu \,(w \mid v_n)_X\, v_n \\ &= \sum_{n\in\mathbb{N}} r_\alpha(\sigma_n^2)\,(w \mid v_n)_X\, |K|^\nu v_n \\ &= |K|^\nu \sum_{n\in\mathbb{N}} r_\alpha(\sigma_n^2)\,(w \mid v_n)_X\, v_n =: |K|^\nu \xi. \end{aligned}$$

The interpolation inequality (3.18) for $r = \nu$ and $s = \nu + 1$ then yields that

$$\|x_\alpha - x^\dagger\|_X = \||K|^\nu \xi\|_X \le \||K|^{\nu+1}\xi\|_X^{\frac{\nu}{\nu+1}} \|\xi\|_X^{\frac{1}{\nu+1}}.$$

Again we estimate the terms separately: For the second factor, we obtain from the definition of ξ, the Bessel inequality, the boundedness of r_α, and the source condition that

$$\|\xi\|_X^2 = \sum_{n\in\mathbb{N}} |r_\alpha(\sigma_n^2)|^2 |\,(w \mid v_n)_X\,|^2 \le C_r^2 \|w\|_X^2 \le C_r^2 \rho^2.$$

For the first factor, we use Lemma 3.13 (i), (iii), $Kx^\dagger = y$ since $y \in \mathcal{R}(K)$, and the productive zero to obtain

$$\begin{aligned} \||K|^{\nu+1}\xi\|_X &= \||K|(|K|^\nu \xi)\|_X = \||K|(x_\alpha - x^\dagger)\|_X = \|K(x_\alpha - x^\dagger)\|_Y = \|Kx_\alpha - y\|_Y \\ &\le \|Kx_\alpha^\delta - y^\delta\|_Y + \|y - y^\delta - K(x_\alpha - x_\alpha^\delta)\|_Y. \end{aligned}$$

Yet again we estimate the terms separately: First, by the choice $\alpha(\delta, y^\delta)$ according to the discrepancy principle we have that $\|Kx_\alpha^\delta - y^\delta\|_Y \le \tau\delta$. For the second term, we write

$$y - Kx_\alpha = y - KR_\alpha y = (\operatorname{Id} - K\varphi_\alpha(K^*K)K^*)y$$

and analogously for $y^\delta - Kx_\alpha^\delta$. Hence,

$$\begin{aligned} \|y - y^\delta - K(x_\alpha - x_\alpha^\delta)\|_Y^2 &= \|(\operatorname{Id} - K\varphi_\alpha(K^*K)K^*)(y - y^\delta)\|_Y^2 \\ &= \sum_{n\in\mathbb{N}} \left|r_\alpha(\sigma_n^2)\right|^2 \left|\left(y - y^\delta \,\middle|\, u_n\right)_Y\right|^2 \\ &\le C_r^2 \delta^2, \end{aligned} \tag{5.13}$$

where we have used for the second equality that (compare (5.3))

$$K\varphi_\alpha(K^*K)K^*(y-y^\delta) = \sum_{n\in\mathbb{N}} \varphi_\alpha(\sigma_n^2)\sigma_n^2 \left(y-y^\delta \,\middle|\, u_n\right)_Y u_n.$$

Together, we obtain for the first term in (5.12) that

$$\|x_\alpha - x^\dagger\|_X \le (\tau + C_r)^{\frac{\nu}{\nu+1}}\delta^{\frac{\nu}{\nu+1}} C_r^{\frac{1}{\nu+1}}\rho^{\frac{1}{\nu+1}} =: C_1\delta^{\frac{\nu}{\nu+1}}\rho^{\frac{1}{\nu+1}}.$$

It remains to estimate the second term (5.12). For this, we use Lemma 5.4 and the condition (5.7) to obtain

$$\begin{aligned}(5.14)\qquad \|x_\alpha^\delta - x_\alpha\|_X \le \|R_\alpha\|_{\mathbb{L}(Y,X)}\delta &\le \sqrt{C_\varphi} \sup_{\lambda\in(0,\kappa]} \sqrt{|\varphi_\alpha(\lambda)|}\delta \\ &\le C_\varphi \alpha(\delta, y^\delta)^{-1/2}\delta.\end{aligned}$$

To show that the right-hand side is of the optimal order, we need to bound $\alpha(\delta, y^\delta)$ in terms of δ appropriately. First, its choice according to the discrepancy principle implies in particular that

$$\tau\delta < \|Kx_{2\alpha}^\delta - y^\delta\|_Y \le \|Kx_{2\alpha} - y\|_Y + \|y - y^\delta - K(x_{2\alpha} - x_{2\alpha}^\delta)\|_Y$$

(where the choice $2\alpha > \alpha$ was arbitrary and for the sake of simplicity). Since the estimate (5.13) is uniform in $\alpha > 0$, we also have that

$$\|y - y^\delta - K(x_{2\alpha} - x_{2\alpha}^\delta)\|_Y \le C_r\delta$$

and thus that

$$\|Kx_{2\alpha} - y\|_Y > \tau\delta - \|y - y^\delta - K(x_{2\alpha} - x_{2\alpha}^\delta)\|_Y \ge (\tau - C_r)\delta.$$

Conversely, we obtain from Lemma 5.8 and condition (5.8) for $\nu + 1 \le \nu_0$ the estimate

$$\|Kx_{2\alpha} - y\|_Y \le \omega_{\nu+1}(2\alpha(\delta, y^\delta))\rho \le C_{\nu+1}(2\alpha(\delta, y^\delta))^{\frac{\nu+1}{2}}\rho.$$

Since $\tau > C_r$ by assumption, this implies that

$$\delta \le (\tau - C_r)^{-1}C_{\nu+1}2^{\frac{\nu+1}{2}}\alpha(\delta, y^\delta)^{\frac{\nu+1}{2}}\rho =: C_\tau\alpha(\delta, y^\delta)^{\frac{\nu+1}{2}}\rho,$$

i.e.,

$$(5.15)\qquad \alpha(\delta, y^\delta)^{-1/2} \le C_\tau^{\frac{1}{\nu+1}}\delta^{-\frac{1}{\nu+1}}\rho^{\frac{1}{\nu+1}}.$$

Inserting this into (5.14) now yields

$$\|x_\alpha^\delta - x_\alpha\|_X \le C_\varphi C_\tau^{\frac{1}{\nu+1}}\delta\delta^{-\frac{1}{\nu+1}}\rho^{\frac{1}{\nu+1}} =: C_2\delta^{\frac{\nu}{\nu+1}}\rho^{\frac{1}{\nu+1}}.$$

Combining the estimates for the two terms in (5.12), we obtain that

$$\|x_\alpha^\delta - x^\dagger\|_X \le (C_1 + C_2)\delta^{\frac{\nu}{\nu+1}}\rho^{\frac{1}{\nu+1}}$$

and thus the order optimality. Theorem 4.14 for $\nu = \nu_0 - 1$ and $\tau_0 = C_r$ then shows that R_α together with the discrepancy principle as parameter choice rule $\alpha_\tau = \alpha(\tau\delta, y^\delta)$ for all $\tau > C_r$ is a regularization method. □

Example 5.12 (truncated singular value decomposition). We have

$$|r_\alpha(\lambda)| = \begin{cases} 1 - \lambda\frac{1}{\lambda} = 0 & \lambda \geq \alpha \\ 1 & \lambda < \alpha \end{cases}$$

and hence $C_r = 1$. Since the truncated singular value decomposition has infinite qualification, it is also an order optimal regularization method for any $\nu > 0$ when combined with the discrepancy principle for arbitrary $\tau > 1$.

If a filter only has finite qualification, the Morozov discrepancy principle only leads to an order optimal regularization method for $\nu > \nu_0 - 1$; this is the price to pay for the indirect control of $\alpha(\delta, y^\delta)$ through the residual (cf. (5.5)). However, there are improved discrepancy principles that measure the residual in adapted norms and thus lead to order optimal regularization methods also for $\nu \in (\nu_0 - 1, \nu_0]$; see, e.g., [Engl, Hanke & Neubauer 1996, Chapter 4.4].

HEURISTIC CHOICE RULES

We consider as an example the Hanke–Raus rule: Define for $y^\delta \in Y$ the function

$$\Psi : (0, \kappa] \to \mathbb{R}, \qquad \Psi(\alpha) = \frac{\|Kx_\alpha^\delta - y^\delta\|_Y}{\sqrt{\alpha}},$$

and choose

$$\alpha(y^\delta) \in \arg\min_{\alpha\in(0,\kappa]} \Psi(\alpha). \tag{5.16}$$

We assume in the following that $y \in \mathcal{R}(K)$ and $\|y\|_Y > \delta$. First, we show a conditional error estimate.

Theorem 5.13. *Let $\{\varphi_\alpha\}_{\alpha>0}$ be a filter with qualification $\nu_0 > 0$, i.e., satisfying (5.7) as well as (5.8) for all $\nu \in (0, \nu_0]$. Furthermore, assume there exists a minimizer $\alpha^* := \alpha(y^\delta) \in (0, \kappa]$ of Ψ with*

$$\delta^* := \|Kx_{\alpha^*}^\delta - y^\delta\|_Y > 0. \tag{5.17}$$

Then there exists a $c > 0$ such that for all $\nu \in (0, \nu_0 - 1]$ and $\rho \geq 0$ with $x^\dagger \in X_{\nu,\rho}$,

$$\|x_{\alpha^*}^\delta - x^\dagger\|_X \leq c\left(1 + \frac{\delta}{\delta^*}\right)\max\{\delta, \delta^*\}^{\frac{\nu}{\nu+1}}\rho^{\frac{1}{\nu+1}}.$$

Proof. Once more we start from the error decomposition

$$\|x_{\alpha^*}^\delta - x^\dagger\|_X \leq \|x_{\alpha^*} - x^\dagger\|_X + \|x_{\alpha^*}^\delta - x_{\alpha^*}\|_X.$$

For the first term, we argue as in the proof of Theorem 5.11 using (5.17) in place of the discrepancy principle to show that

$$\|x_{\alpha^*} - x^\dagger\|_X \leq C_r^{\frac{1}{\nu+1}}(\delta^* + C_r\delta)^{\frac{\nu}{\nu+1}}\rho^{\frac{1}{\nu+1}} \leq C_1 \max\{\delta, \delta^*\}^{\frac{\nu}{\nu+1}}\rho^{\frac{1}{\nu+1}}. \tag{5.18}$$

for some constant $C_1 > 0$.

For the second term, we obtain similarly as for (5.14) using (5.17) (in the form of the productive $1 = \delta^*/\delta^*$) that

$$\|x_{\alpha^*}^\delta - x_{\alpha^*}\|_X \leq C_\varphi \frac{1}{\sqrt{\alpha^*}}\delta = C_\varphi \frac{\delta}{\delta^*}\frac{\|Kx_{\alpha^*}^\delta - y^\delta\|_Y}{\sqrt{\alpha^*}} = C_\varphi \frac{\delta}{\delta^*}\Psi(\alpha^*).$$

Again, we need to bound the last factor by the correct power of δ, for which we use the choice rule. In this case, (5.16) states that $\Psi(\alpha^*) \leq \Psi(\alpha)$ for all $\alpha \in (0, \kappa]$. The idea is now to compare with α chosen according to the discrepancy principle, which however need not be feasible (it may be larger than κ). Let therefore $\bar\alpha := \alpha(\delta, y^\delta)$ be chosen such that (5.10) holds. If $\bar\alpha \leq \kappa$, then (5.15) yields that

$$\Psi(\alpha^*) \leq \Psi(\bar\alpha) \leq (\tau\delta)(C_\tau^{\frac{1}{\nu+1}}\delta^{-\frac{1}{\nu+1}}\rho^{\frac{1}{\nu+1}}) = C_\tau^{\frac{1}{\nu+1}}\tau\delta^{\frac{\nu}{\nu+1}}\rho^{\frac{1}{\nu+1}}. \tag{5.19}$$

On the other hand, if $\bar\alpha > \kappa = \|K\|^2_{\mathbb{L}(X,Y)}$, then by assumption $\|Kx_\kappa^\delta - y^\delta\|_Y \leq \tau\delta$ as well. From

$$\delta < \|y\|_Y = \|Kx^\dagger\|_Y = \|K|K|^\nu w\|_X \leq \|K\|^{\nu+1}_{\mathbb{L}(X,Y)}\rho$$

it then follows that

$$\Psi(\alpha^*) \leq \Psi(\kappa) \leq \tau\delta\|K\|^{-1}_{\mathbb{L}(X,Y)} < \tau\delta\left(\delta^{-\frac{1}{\nu+1}}\rho^{\frac{1}{\nu+1}}\right) = \tau\delta^{\frac{\nu}{\nu+1}}\rho^{\frac{1}{\nu+1}}. \tag{5.20}$$

In both cases, we thus obtain that

$$\|x_{\alpha^*}^\delta - x_{\alpha^*}\|_X \leq C_2\frac{\delta}{\delta^*}\delta^{\frac{\nu}{\nu+1}}\rho^{\frac{1}{\nu+1}}$$

for some constant $C_2 > 0$. Together with (5.18), this shows the claimed estimate. □

Hence the Hanke–Raus rule would be order optimal if $\delta^* \approx \delta$. Conversely, the rule would fail if $\alpha^* = 0$ or $\delta^* = 0$ occurred. In the later case, $y^\delta \in \mathcal{R}(K)$, and the unboundedness of $K^\dagger$ would imply that $\|K^\dagger y^\delta - K^\dagger y\|_Y$ could be arbitrarily large. We thus need to exclude this case in order to show error estimates. For example, we can assume that there exists an $\varepsilon > 0$ such that

$$y^\delta \in \mathcal{N}_\varepsilon := \left\{y + \eta \in Y \;\middle|\; \|(\mathrm{Id} - P_{\overline{\mathcal{R}}})\eta\|_Y \geq \varepsilon\|\eta\|_Y\right\}, \tag{5.21}$$

where $P_{\overline{\mathcal{R}}}$ denotes the orthogonal projection onto $\overline{\mathcal{R}(K)}$. Intuitively, this means that the noisy data y^δ cannot be arbitrarily close to $\overline{\mathcal{R}(K)}$. Restricted to such data, the Hanke–Raus rule indeed leads to a convergent regularization method.

Theorem 5.14. *Let $\{\varphi_\alpha\}_{\alpha>0}$ be a filter with qualification $\nu_0 > 0$ satisfying (5.7) as well as (5.8). Assume further that (5.21) holds. Then for every $x^\dagger \in X_{\nu,\rho}$ with $\nu \in (0, \nu_0 - 1]$ and $\rho > 0$ and $y = Kx^\dagger$,*

$$\lim_{\delta\to 0} \sup_{y^\delta \in B_\delta(y)\cap \mathcal{N}_\varepsilon} \|x^\delta_{\alpha^*} - x^\dagger\|_X = 0.$$

Proof. Let $y \in \mathcal{R}(K)$ and $y^\delta \in \mathcal{N}_\varepsilon$ with $\|y^\delta - y\|_Y = \delta$. Since $\operatorname{Id} - P_{\overline{\mathcal{R}}}$ is an orthogonal projection and therefore has operator norm 1, we have for all $\alpha > 0$ that

$$\begin{aligned} \|Kx^\delta_\alpha - y^\delta\|_Y &\geq \|(\operatorname{Id} - P_{\overline{\mathcal{R}}})(Kx^\delta_\alpha - y^\delta)\|_Y = \|(\operatorname{Id} - P_{\overline{\mathcal{R}}})y^\delta\|_Y \\ &= \|(\operatorname{Id} - P_{\overline{\mathcal{R}}})(y^\delta - y)\|_Y \geq \varepsilon\|y^\delta - y\|_Y \\ &= \varepsilon\delta > 0. \end{aligned} \tag{5.22}$$

This implies that the numerator of $\Psi(\alpha)$ is bounded from below, and hence $\Psi(\alpha) \to \infty$ for $\alpha \to 0$. The infimum over all $(0, \kappa]$ therefore must be attained for $\alpha^* > 0$. In particular, it follows from (5.22) that

$$\delta^* = \|Kx^\delta_{\alpha^*} - y^\delta\|_Y \geq \varepsilon\delta > 0.$$

We thus obtain from Theorem 5.13 and the estimate $\delta \leq \varepsilon^{-1}\delta^*$ that

$$\|x^\delta_{\alpha^*} - x^\dagger\|_X \leq C_\varepsilon (\delta^*)^{\frac{\nu}{\nu+1}} \rho^{\frac{1}{\nu+1}}$$

for some constant $C_\varepsilon > 0$. It thus suffices to show that $\delta \to 0$ implies that $\delta^* \to 0$ as well. But this follows from $\alpha^* \leq \kappa$ and (5.19) or (5.20), since as $\delta \to 0$, we have that

$$\delta^* = \|Kx^\delta_{\alpha^*} - y^\delta\|_Y = \sqrt{\alpha^*}\Psi(\alpha^*) \leq \sqrt{\kappa}\Psi(\alpha^*) \leq \sqrt{\kappa} \max\{1, C_\tau^{\frac{1}{\nu+1}}\}\tau\delta^{\frac{\nu}{\nu+1}}\rho^{\frac{1}{\nu+1}} \to 0. \qquad \square$$

Under similar assumptions (and with more effort), it is also possible to show order optimality of the Hanke–Raus rule as well as of related minimization-based heuristic choice rules; see [Kindermann 2011].

6 TIKHONOV REGULARIZATION

Due to its central role in the theory and practice of inverse problems, we again consider in more detail Tikhonov regularization, which corresponds to the filter

$$\varphi_\alpha(\lambda) = \frac{1}{\lambda + \alpha}.$$

We get to the point quickly since we are well prepared. As we have already noted in Example 5.2 (ii), the filter φ_α is continuous, converges to $\frac{1}{\lambda}$ as $\alpha \to 0$, is uniformly bounded by α^{-1}, and satisfies

$$\lambda\varphi_\alpha(\lambda) = \frac{\lambda}{\lambda + \alpha} < 1 =: C_\varphi \qquad \text{for all } \alpha > 0.$$

By Theorem 5.6, the operator $R_\alpha = \varphi_\alpha(K^*K)K^*$ is therefore a regularization, satisfies by Lemma 5.4

$$\|R_\alpha\|_{\mathbb{L}(Y,X)} \le \frac{1}{\sqrt{\alpha}},$$

and by Theorem 4.6 leads together with the a priori choice rule $\alpha(\delta) = \delta$ to a convergent regularization method.

To show convergence rates, we apply Theorem 5.9 (for a priori choice rules) and Theorem 5.11 (for the Morozov discrepancy principle). First, since $\varphi_\alpha(\lambda) \le \alpha^{-1} = C_\varphi\alpha^{-1}$ for all $\alpha > 0$, the condition (5.7) is satisfied. Furthermore,

$$r_\alpha(\lambda) = 1 - \lambda\varphi_\alpha(\lambda) = \frac{\alpha}{\lambda + \alpha} \le 1 =: C_r \qquad \text{for all } \alpha > 0, \lambda \in (0, \kappa].$$

To show the second condition (5.8), we have to estimate

$$\omega_\nu(\alpha) = \sup_{\lambda \in (0,\kappa]} \lambda^{\nu/2}|r_\alpha(\lambda)| = \sup_{\lambda \in (0,\kappa]} \frac{\lambda^{\nu/2}\alpha}{\lambda + \alpha} =: \sup_{\lambda \in (0,\kappa]} h_\alpha(\lambda)$$

by $C_\nu\alpha^{\nu/2}$ for a constant $C_\nu > 0$. To do this, we consider $h_\alpha(\lambda)$ for fixed $\alpha > 0$ as a function of λ and compute

$$h_\alpha'(\lambda) = \frac{\alpha\frac{\nu}{2}\lambda^{\nu/2-1}(\lambda + \alpha) - \alpha\lambda^{\nu/2}}{(\lambda + \alpha)^2} = \frac{\alpha\lambda^{\nu/2-1}}{(\lambda + \alpha)^2}\left(\frac{\nu}{2}\alpha + \left(\frac{\nu}{2} - 1\right)\lambda\right).$$

For $\nu \geq 2$, the function $h_\alpha(\lambda)$ is therefore increasing, and the maximum over all $\lambda \in (0, \kappa]$ is attained in $\lambda^* := \kappa$. In this case,

$$\omega_\nu(\alpha) = h_\alpha(\kappa) = \frac{\alpha\kappa^{\nu/2}}{\kappa + \alpha} \leq \kappa^{\nu/2-1}\alpha.$$

We thus obtain the desired estimate (only) for $\nu = 2$.

For $\nu \in (0, 2)$, we can compute the root of $h'_\alpha(\lambda)$ as $\lambda^* := \frac{\alpha\frac{\nu}{2}}{1-\frac{\nu}{2}}$. There, $h''_\alpha(\lambda^*) < 0$, which yields for all $\alpha > 0$ that

$$\omega_\nu(\alpha) = h_\alpha(\lambda^*) = \frac{\alpha\left(\alpha\frac{\nu}{2}\left(1-\frac{\nu}{2}\right)^{-1}\right)^{\nu/2}}{\alpha + \alpha\frac{\nu}{2}\left(1-\frac{\nu}{2}\right)^{-1}} \leq \left(\frac{\nu}{2}\left(1-\frac{\nu}{2}\right)^{-1}\right)^{\nu/2}\alpha^{\nu/2}$$

and hence the desired estimate.

Tikhonov regularization thus has at least (and, as we will show, at most) qualification $\nu_0 = 2$. The corresponding order optimality for a priori and a posteriori choice rules now follows easily from Theorem 5.9 and Theorem 5.11, respectively

Corollary 6.1. *For all $\nu \in (0, 2]$, Tikhonov regularization together with the parameter choice rule*

$$c\left(\frac{\delta}{\rho}\right)^{\frac{2}{\nu+1}} \leq \alpha(\delta) \leq C\left(\frac{\delta}{\rho}\right)^{\frac{2}{\nu+1}} \qquad \text{for } C > c > 0$$

is an order optimal regularization method. In particular, for $\alpha \sim \delta^{2/3}$,

$$\|x^\delta_{\alpha(\delta)} - x^\dagger\|_X \leq c\delta^{\frac{2}{3}} \qquad \text{for all } x^\dagger \in \mathcal{R}(K^*K) \text{ and } y^\delta \in B_\delta(Kx^\dagger).$$

Corollary 6.2. *For all $\nu \in (0, 1]$ and $\tau > 1$, Tikhonov regularization together with the parameter choice rule*

$$\|Kx^\delta_{\alpha(\delta,y^\delta)} - y^\delta\|_Y \leq \tau\delta < \|Kx^\delta_\alpha - y^\delta\|_Y \qquad \text{for all } \alpha > \alpha(\delta, y^\delta)$$

is an order optimal regularization method. In particular,

$$\|x^\delta_{\alpha(\delta,y^\delta)} - x^\dagger\|_X \leq c\delta^{\frac{1}{2}} \qquad \text{for all } x^\dagger \in \mathcal{R}(K^*) \text{ and } y^\delta \in B_\delta(Kx^\dagger).$$

In fact, the qualification cannot be larger than 2; Tikhonov regularization thus *saturates* in contrast to, e.g., the truncated singular value decomposition. To show this, we first derive the alternative characterization that was promised in Example 5.2 (ii).

Lemma 6.3. *Let $y \in Y$ and $\alpha > 0$. Then $x = x_\alpha := R_\alpha y$ if and only if*

$$(K^*K + \alpha \operatorname{Id})x_\alpha = K^*y. \tag{6.1}$$

Proof. We use the singular system $\{(\sigma_n, u_n, v_n)\}_{n\in\mathbb{N}}$ of K to obtain

$$\alpha x_\alpha = \sum_{n\in\mathbb{N}} \alpha \frac{\sigma_n}{\sigma_n^2+\alpha} (y \,|\, u_n)_Y \, v_n$$

as well as

$$\begin{aligned} K^*Kx_\alpha &= \sum_{n\in\mathbb{N}} \frac{\sigma_n}{\sigma_n^2+\alpha} (y \,|\, u_n)_Y \, K^*Kv_n \\ &= \sum_{n\in\mathbb{N}} \sigma_n^2 \frac{\sigma_n}{\sigma_n^2+\alpha} (y \,|\, u_n)_Y \, v_n. \end{aligned}$$

This implies that

$$(K^*K + \alpha \operatorname{Id})x_\alpha = \sum_{n\in\mathbb{N}} \sigma_n \, (y \,|\, u_n)_Y \, v_n = K^*y.$$

Conversely, let $x \in X$ be a solution of (6.1). Inserting the representation

$$x = \sum_{n\in\mathbb{N}} (x \,|\, v_n)_X \, v_n + P_{\mathcal{N}}x \tag{6.2}$$

into (6.1) then yields

$$\sum_{n\in\mathbb{N}} (\sigma_n^2+\alpha) \, (x \,|\, v_n)_X \, v_n + \alpha P_{\mathcal{N}}x = (K^*K + \alpha \operatorname{Id})x = K^*y = \sum_{n\in\mathbb{N}} \sigma_n \, (y \,|\, u_n)_Y \, v_n.$$

Since $\{v_n\}_{n\in\mathbb{N}}$ is an orthonormal basis of $\overline{\mathcal{R}(K^*)} = \mathcal{N}(K)^\perp$, we must have $P_{\mathcal{N}}x = 0$. Equating coefficients then shows that

$$(x \,|\, v_n)_X = \frac{\sigma_n}{\sigma_n^2+\alpha} (y \,|\, u_n)_Y \qquad \text{for all } n \in \mathbb{N}.$$

Inserting this into (6.2) in turn yields

$$x = \sum_{n\in\mathbb{N}} (x \,|\, v_n)_X \, v_n = \sum_{n\in\mathbb{N}} \frac{\sigma_n}{\sigma_n^2+\alpha} (y \,|\, u_n)_Y \, v_n = x_\alpha,$$

i.e., x_α is the unique solution of (6.1). □

The practical value of the characterization (6.1) cannot be emphasized enough: Instead of a singular value decomposition, it suffices to compute the solution of a *well-posed* linear equation (for a selfadjoint positive definite operator), which can be done using standard methods.

We now show that in general there cannot be an a priori choice rule for which the regularization error $\|x^\delta_{\alpha(\delta)} - x^\dagger\|_X$ tends to zero faster than $\delta^{2/3}$.

Theorem 6.4. *Let $K \in \mathcal{K}(X,Y)$ have infinite-dimensional range and let $y \in \mathcal{R}(K)$. If there exists an a priori parameter choice rule α with $\lim_{\delta\to 0}\alpha(\delta) = 0$ such that*

$$\lim_{\delta\to 0}\sup_{y^\delta\in B_\delta(y)} \|x^\delta_{\alpha(\delta)} - x^\dagger\|_X \delta^{-\frac{2}{3}} = 0, \tag{6.3}$$

then $x^\dagger = 0$.

Proof. Assume to the contrary that $x^\dagger \neq 0$. We first show that the given assumptions imply that $\alpha(\delta)\delta^{-2/3} \to 0$. For this, we use the characterization (6.1) for x_α^δ and y^δ to write

$$(K^*K + \alpha(\delta)\,\mathrm{Id})\left(x^\delta_{\alpha(\delta)} - x^\dagger\right) = K^*y^\delta - K^*y - \alpha(\delta)x^\dagger.$$

Together with $\kappa = \|K^*K\|_{\mathbb{L}(X,X)} = \|K^*\|^2_{\mathbb{L}(Y,X)}$, this implies that

$$|\alpha(\delta)|\|x^\dagger\|_X \leq \sqrt{\kappa}\delta + (\alpha(\delta)+\kappa)\|x^\delta_{\alpha(\delta)} - x^\dagger\|_X.$$

Multiplying this with $\delta^{-2/3}$ and using the assumption (6.3) as well as $x^\dagger \neq 0$ then yields that

$$|\alpha(\delta)|\delta^{-2/3} \leq \|x^\dagger\|_X^{-1}\left(\sqrt{\kappa}\delta^{\frac{1}{3}} + (\alpha(\delta)+\kappa)\|x^\delta_{\alpha(\delta)} - x^\dagger\|_X\delta^{-\frac{2}{3}}\right) \to 0.$$

We now construct a contradiction. Let $\{(\sigma_n, u_n, v_n)\}_{n\in\mathbb{N}}$ be a singular system of K and define

$$\delta_n := \sigma_n^3 \quad\text{and}\quad y_n := y + \delta_n u_n, \qquad n\in\mathbb{N},$$

such that $\|y_n - y\|_Y = \delta_n \to 0$ as $n\to\infty$. Furthermore, setting $\alpha_n := \alpha(\delta_n)$, we have that

$$\begin{aligned}
x^{\delta_n}_{\alpha_n} - x^\dagger &= (x^{\delta_n}_{\alpha_n} - x_{\alpha_n}) + (x_{\alpha_n} - x^\dagger)\\
&= R_\alpha(y_n - y) + (x_{\alpha_n} - x^\dagger)\\
&= \sum_{m\in\mathbb{N}} \frac{\sigma_m}{\sigma_m^2 + \alpha_n}(\delta_n u_n \mid u_m)_Y\, v_m + (x_{\alpha_n} - x^\dagger)\\
&= \frac{\delta_n\sigma_n}{\sigma_n^2+\alpha_n}v_n + (x_{\alpha_n} - x^\dagger).
\end{aligned}$$

Together with the assumption (6.3) for $y^\delta = y_n$ as well as for $y^\delta = y$, this implies that

$$\frac{\sigma_n\delta_n^{1/3}}{\sigma_n^2+\alpha_n} \leq \|x^{\delta_n}_{\alpha_n} - x^\dagger\|_X\delta_n^{-2/3} + \|x_{\alpha_n} - x^\dagger\|_X\delta_n^{-2/3} \to 0 \qquad \text{as } n\to\infty.$$

On the other hand, $\sigma_n = \delta_n^{1/3}$ and $\alpha_n\delta_n^{-2/3} \to 0$ imply that

$$\frac{\sigma_n\delta_n^{1/3}}{\sigma_n^2+\alpha_n} = \frac{\delta_n^{2/3}}{\delta_n^{2/3}+\alpha_n} = \frac{1}{1+\alpha_n\delta_n^{-2/3}} \to 1 \qquad \text{as } n\to\infty$$

and hence the desired contradiction. ∎

Comparing the characterization (6.1) with the normal equations (3.6) suggest that Tikhonov regularization also has a minimization property. This is indeed the case.

Theorem 6.5. *Let $y \in Y$ and $\alpha > 0$. Then $x_\alpha := R_\alpha y$ is the unique minimizer of the* Tikhonov functional

$$J_\alpha(x) := \frac{1}{2}\|Kx - y\|_Y^2 + \frac{\alpha}{2}\|x\|_X^2. \tag{6.4}$$

Proof. A minimizer $\bar{x} \in X$ of J_α is defined as satisfying $J_\alpha(\bar{x}) \leq J_\alpha(x)$ for all $x \in X$. We therefore take the difference of functional values for arbitrary $x \in X$ and for the solution x_α of (6.1) and rearrange the inner products to obtain

$$\begin{aligned}
J_\alpha(x) - J_\alpha(x_\alpha) &= \frac{1}{2}\left(Kx - y \,|\, Kx - y\right)_Y + \frac{\alpha}{2}\left(x \,|\, x\right)_X \\
&\quad - \frac{1}{2}\left(Kx_\alpha - y \,|\, Kx_\alpha - y\right)_Y - \frac{\alpha}{2}\left(x_\alpha \,|\, x_\alpha\right)_X \\
&= \frac{1}{2}\|Kx - Kx_\alpha\|_Y^2 + \frac{\alpha}{2}\|x - x_\alpha\|_X^2 + \left(K^*(Kx_\alpha - y) + \alpha x_\alpha \,|\, x - x_\alpha\right)_X \\
&= \frac{1}{2}\|Kx - Kx_\alpha\|_Y^2 + \frac{\alpha}{2}\|x - x_\alpha\|_X^2 \\
&\geq 0,
\end{aligned}$$

where we have used (6.1) in the last equality. Hence, x_α is a minimizer of J_α.

Conversely, if $J_\alpha(x) - J_\alpha(\bar{x}) \geq 0$ for all $x \in X$, we in particular have for $x = \bar{x} + tz$ with arbitrary $t > 0$ and $z \in X$ that

$$0 \leq J_\alpha(\bar{x} + tz) - J_\alpha(\bar{x}) = \frac{t^2}{2}\|Kz\|_Y^2 + \frac{t^2\alpha}{2}\|z\|_X^2 + t\left(K^*(K\bar{x} - y) + \alpha\bar{x} \,|\, z\right)_X.$$

Dividing by $t > 0$ and passing to the limit $t \to 0$ then yields

$$\left(K^*(K\bar{x} - y) + \alpha\bar{x} \,|\, z\right)_X \geq 0.$$

Since $z \in X$ was arbitrary, this can only hold if $K^*K\bar{x} + \alpha\bar{x} = K^*y$. As x_α is the unique solution of (6.1), we obtain $\bar{x} = x_\alpha$. Hence, x_α is in fact the unique minimizer of (6.4). □

The characterization of Tikhonov regularization as minimization of the functional (6.4) furthermore yields another connection to the minimum norm solution $x^\dagger$: Instead of insisting on a least squares solution, whose norm need not be bounded for $y \notin \mathcal{D}(K^\dagger)$, we look for an approximation that minimizes (squared) residual norm $\|Kx - y\|_Y^2$ *together* with the (squared) norm $\|x\|_X^2$.[1] Here the regularization parameter α determines the trade-off:

[1]This is also the form in which this regularization was introduced by Andreĭ Nikolaevich Tikhonov, a prominent Russian mathematician of the 20th century; see [Tikhonov 1963a; Tikhonov 1963b].

the smaller the noise level δ, the more importance one can put on the minimization of the residual (i.e., the smaller α can be chosen). Conversely, a larger noise level requires putting more weight on minimizing the *penalty term* $\|x\|_X^2$ (and hence choosing a larger α) in order to obtain a stable approximation.

In addition, this characterization can be used to derive monotonicity properties of the *value functions*

$$f(\alpha) := \frac{1}{2}\|Kx_\alpha^\delta - y^\delta\|_Y^2, \qquad g(\alpha) := \frac{1}{2}\|x_\alpha^\delta\|_X^2,$$

and

$$j(\alpha) := J_\alpha(x_\alpha^\delta) = f(\alpha) + \alpha g(\alpha) = J_\alpha(x_\alpha^\delta).$$

Lemma 6.6. *The value functions f and g are monotone in the sense that for all $\alpha_1, \alpha_2 > 0$,*

$$(f(\alpha_1) - f(\alpha_2))\,(\alpha_1 - \alpha_2) \geq 0, \tag{6.5}$$

$$(g(\alpha_1) - g(\alpha_2))\,(\alpha_1 - \alpha_2) \leq 0. \tag{6.6}$$

Proof. The minimization property of $x_{\alpha_1}^\delta$ for J_{α_1} and of $x_{\alpha_2}^\delta$ for J_{α_2} imply that

$$\begin{aligned} f(\alpha_1) + \alpha_1 g(\alpha_1) &\leq f(\alpha_2) + \alpha_1 g(\alpha_2), \\ f(\alpha_2) + \alpha_2 g(\alpha_2) &\leq f(\alpha_1) + \alpha_2 g(\alpha_1). \end{aligned}$$

Adding these inequalities and rearranging immediately yields (6.6). Dividing the first inequality by $\alpha_1 > 0$, the second by $\alpha_2 > 0$, and adding both yields

$$\frac{1}{\alpha_1}\left(f(\alpha_1) - f(\alpha_2)\right) \leq \frac{1}{\alpha_2}\left(f(\alpha_1) - f(\alpha_2)\right).$$

Multiplying by $\alpha_1\alpha_2 > 0$ and rearranging then yields (6.5). □

As expected, the residual norm is decreasing and the norm of x_α^δ is increasing as $\alpha \to 0$. We next consider for the value function j the one-sided difference quotients

$$\begin{aligned} D^+ j(\alpha) &:= \lim_{t\to 0^+} \frac{j(\alpha + t) - j(\alpha)}{t}, \\ D^- j(\alpha) &:= \lim_{t\to 0^-} \frac{j(\alpha + t) - j(\alpha)}{t}. \end{aligned}$$

Lemma 6.7. *For all $\alpha > 0$,*

$$\begin{aligned} D^+ j(\alpha) \quad &\leq \quad g(\alpha) \quad \leq \quad D^- j(\alpha), \\ j(\alpha) - \alpha D^- j(\alpha) \quad &\leq \quad f(\alpha) \quad \leq \quad j(\alpha) - \alpha D^+(\alpha). \end{aligned}$$

Proof. For any $\alpha, \tilde\alpha > 0$, the minimization property for j yields that

$$j(\tilde\alpha) = f(\tilde\alpha) + \tilde\alpha g(\tilde\alpha) \le f(\alpha) + \tilde\alpha g(\alpha).$$

Hence,

$$\begin{aligned} j(\alpha) - j(\tilde\alpha) &= f(\alpha) + \alpha g(\alpha) - f(\tilde\alpha) - \tilde\alpha g(\tilde\alpha) \\ &\ge f(\alpha) + \alpha g(\alpha) - f(\alpha) - \tilde\alpha g(\alpha) \\ &= (\alpha - \tilde\alpha) g(\alpha), \end{aligned}$$

which implies for $\tilde\alpha := \alpha + t > \alpha$ with $t > 0$ that

$$\frac{j(\alpha + t) - j(\alpha)}{t} \le g(\alpha).$$

Passing to the limit $t \to 0$ thus shows that $D^+ j(\alpha) \le g(\alpha)$. The corresponding inequality for $D^- j(\alpha)$ follows analogously with $t < 0$.

The remaining inequalities follow from this together with the definition of j; for example, using

$$j(\alpha) = f(\alpha) + \alpha g(\alpha) \le f(\alpha) + \alpha D^- j(\alpha),$$

and rearranging. □

By one of Lebesgue's theorems (see [Hewitt & Stromberg 1975, Theorem V.17.12]), a monotone function is differentiable almost everywhere (i.e., $D^- f \ne D^+ f$ on at most a set of Lebesgue measure zero). Hence, f and g and therefore also $j = f + \alpha g$ are differentiable almost everywhere, and we obtain the following expression for the derivative of the latter.

Corollary 6.8. *For almost all $\alpha > 0$, the value function j is differentiable with*

$$j'(\alpha) = g(\alpha).$$

This characterization can be useful for example when implementing minimization-based heuristic parameter choice rules.

Furthermore, Theorem 6.5 suggests a new interpretation of the simplest source condition $x^\dagger \in X_1 = \mathcal{R}(K^*)$. Since the minimizer of (6.4) does not change when dividing the Tikhonov functional by $\alpha > 0$, the minimizer x_α^δ is also a minimizer of

$$\min_{x \in X} \frac{1}{2\alpha} \|Kx - y^\delta\|_Y^2 + \frac{1}{2}\|x\|_X^2. \tag{6.7}$$

Now we want $x_\alpha^\delta \to x^\dagger$ as $\delta \to 0$ and $\alpha \to 0$. Formally passing to the limits in (6.7), i.e., first replacing y^δ with $y \in \mathcal{R}(K)$ and then letting $\alpha \to 0$, we see that the limit functional can only have a finite minimum in some $\bar{x}$ if $K\bar{x} = y$. The limit functional is therefore given by

$$\min_{x \in X,\ Kx=y} \frac{1}{2}\|x\|_X^2. \tag{6.8}$$

We again proceed formally. Introducing the Lagrange multiplier $p \in Y$, we can write (6.8) as the unconstrained saddle-point problem

$$\min_{x\in X} \max_{p\in Y} \frac{1}{2}\|x\|_X^2 - (p \mid Kx - y)_Y .$$

For $(\bar{x}, \bar{p}) \in X \times Y$ to be a saddle point, the partial derivatives with respect to both x and p have to vanish, leading to the conditions

$$\begin{cases} \bar{x} = K^*\bar{p}, \\ K\bar{x} = y. \end{cases}$$

But for $y \in \mathcal{R}(K)$, the solution of (6.8) describes exactly the minimum norm solution $x^\dagger$, i.e., $\bar{x} = x^\dagger$. The existence of a Lagrange multiplier $\bar{p}$ with $x^\dagger = K^*\bar{p}$ is therefore equivalent to the source condition $x^\dagger \in \mathcal{R}(K^*)$. (Since K^* need not be surjective, this is a non-trivial assumption.) Intuitively, this makes sense: If we want to approximate $x^\dagger$ by a sequence of minimizers x_α^δ, the limit $x^\dagger$ should itself be a minimizer (of an appropriate limit problem).

Finally, the interpretation of Tikhonov regularization as minimizing a functional can – in contrast to the construction via the singular value decomposition – be extended to *nonlinear* operator equations as well as to equations in Banach spaces. It can further be generalized by replacing the squared norms by other *discrepancy* and *penalty functionals*. Of course, this also entails generalized source conditions. We will return to this in Chapter 10.

7 LANDWEBER REGULARIZATION

The usual starting point for deriving Landweber regularization is the characterization from Corollary 3.6 of the minimum norm solution as the solution $x \in \mathcal{N}(K)^\perp$ of the normal equations (3.6). These can be written equivalently for any $\omega > 0$ as the fixed point equation

$$x = x - \omega(K^*Kx - K^*y) = x + \omega K^*(y - Kx).$$

The corresponding fixed-point iteration – also known as *Richardson iteration*[1]– is

$$x_n = x_{n-1} + \omega K^*(y - Kx_{n-1}), \qquad n \in \mathbb{N}, \tag{7.1}$$

for some $x_0 \in X$. Here we only consider $x_0 = 0$ for the sake of simplicity. The Banach Fixed-Point Theorem ensures that this iteration converges to a solution of the normal equations if $y \in \mathcal{R}(K)$ and $\|\operatorname{Id} - \omega K^*K\|_{\mathbb{L}(X,X)} < 1$. Since $x_0 = 0 \in \mathcal{R}(K^*)$, an induction argument shows that $x_n \in \mathcal{R}(K^*) \subset \mathcal{N}(K)^\perp$ for all $n \in \mathbb{N}$, and therefore $x_n \to x^\dagger$. If $y^\delta \notin \mathcal{R}(K)$, however, no convergence can be expected. The ideas is therefore to stop the iteration early, i.e., take x_m for an appropriate $m \in \mathbb{N}$ as the regularized approximation. The *stopping index* $m \in \mathbb{N}$ thus plays the role of the regularization parameter here, which fits into the framework of Chapter 5 if we set $\alpha = \frac{1}{m} > 0$.[2]

Performing m steps of the iteration (7.1) can be formulated as a spectral regularization. For this, we first derive a recursion-free characterization of the final iterate x_m.

Lemma 7.1. *If $x_0 = 0$, then*

$$x_m = \omega \sum_{n=0}^{m-1} (\operatorname{Id} - \omega K^*K)^n K^*y \qquad \text{for all } m \in \mathbb{N}.$$

[1]This method for the solution of linear systems of equations traces back to Lewis Fry Richardson. He also proposed in 1922 the modern method of weather prediction by numerical simulation. (His own first attempt in 1910 – by hand! – was correct in principle but gave wrong results due to noisy input data. Weather prediction is an ill-posed problem!)

[2]This method was first proposed for the solution of ill-posed operator equations by Lawrence Landweber. In [Landweber 1951], he shows the convergence for $y \in \mathcal{R}(K)$; otherwise, he then writes, “such a sequence may give useful successive approximations”.

Proof. We proceed by induction. For $m = 1$,

$$x_1 = \omega K^* y = \omega(\mathrm{Id} - \omega K^*K)^0 K^* y.$$

Let now $m \in \mathbb{N}$ be arbitrary, and let the claim hold for x_m. Then

$$\begin{aligned} x_{m+1} &= x_m + \omega K^*(y - Kx_m) \\ &= (\mathrm{Id} - \omega K^*K)x_m + \omega K^* y \\ &= (\mathrm{Id} - \omega K^*K)\left(\omega \sum_{n=0}^{m-1} (\mathrm{Id} - \omega K^*K)^n K^* y\right) + \omega K^* y \\ &= \omega \sum_{n=0}^{m-1} (\mathrm{Id} - \omega K^*K)^{n+1} K^* y + \omega(\mathrm{Id} - \omega K^*K)^0 K^* y \\ &= \omega \sum_{n=0}^{m} (\mathrm{Id} - \omega K^*K)^n K^* y. \end{aligned}$$ □

Performing m steps of the *Landweber iteration* (7.1) is thus equivalent to applying a linear operator, i.e.,

$$x_m = \varphi_m(K^*K)K^* y$$

for

$$\varphi_m(\lambda) = \omega \sum_{n=0}^{m-1} (1 - \omega\lambda)^n = \omega \frac{1 - (1 - \omega\lambda)^m}{1 - (1 - \omega\lambda)} = \frac{1 - (1 - \omega\lambda)^m}{\lambda}.$$

Apart from the notation φ_m instead of φ_α for $\alpha = \frac{1}{m}$ (i.e., considering $m \to \infty$ instead of $\alpha \to 0$), this is exactly the filter from Example 5.2 (iii).

Theorem 7.2. *For any $\omega \in (0, \kappa^{-1})$, the family $\{\varphi_m\}_{m\in\mathbb{N}}$ defines a regularization $\{R_m\}_{m\in\mathbb{N}}$ with $R_m := \varphi_m(K^*K)K^*$.*

Proof. We only have to show that $\varphi_m(\lambda) \to \frac{1}{\lambda}$ as $m \to \infty$ and that $\lambda\varphi_m(\lambda)$ is uniformly bounded for all $m \in \mathbb{N}$. By the assumption on ω, we have $0 < 1 - \omega\lambda < 1$ for all $\lambda \in (0, \kappa]$, which yields $(1 - \omega\lambda)^m \to 0$ as $m \to \infty$ as well as

$$\lambda|\varphi_m(\lambda)| = |1 - (1 - \omega\lambda)^m| \le 1 =: C_\varphi \qquad \text{for all } m \in \mathbb{N} \text{ and } \lambda \in (0, \kappa].$$

Hence $\{\varphi_m\}_{m\in\mathbb{N}}$ is a regularizing filter, and the claim follows from Theorem 5.6. □

Hence the Landweber iteration converges to a minimum norm solution $x^\dagger$ as $m \to \infty$ if and only if $y \in \mathcal{D}(K^\dagger)$; otherwise it diverges. It therefore suggest itself to choose the stopping index by the discrepancy principle: Pick $\tau > 1$ and take $m(\delta, y^\delta)$ such that $x_m^\delta := R_m y^\delta$ satisfies

$$\|Kx^\delta_{m(\delta,y^\delta)} - y^\delta\|_Y \le \tau\delta < \|Kx^\delta_m - y^\delta\|_Y \qquad \text{for all } m < m(\delta, y^\delta). \tag{7.2}$$

(This does not require any additional effort since the residual $y^\delta - Kx_m^\delta$ is computed as part of the iteration (7.1).) The existence of such an $m(\delta, y^\delta)$ is guaranteed by Theorem 4.7 together with Lemma 5.3.

We now address convergence rates, where from now on we assume that $\omega \in (0, \kappa^{-1})$.

Theorem 7.3. *For all $\nu > 0$ and $\tau > 1$, the Landweber iteration (7.1) together with the discrepancy principle (7.2) is an order optimal regularization method.*

Proof. We apply Theorem 5.11, for which we verify the necessary conditions (following the convention $\alpha := \frac{1}{m}$). First, due to $\omega\lambda < 1$ Bernoulli's inequality yields that

$$|\varphi_m(\lambda)| = \frac{|1-(1-\omega\lambda)^m|}{\lambda} \leq \frac{|1-1+m\omega\lambda|}{\lambda} = \omega m = \omega\alpha^{-1} \qquad \text{for all } \lambda \in (0,\kappa]$$

and hence that (5.7) holds. (Clearly for $\omega \leq 1$; otherwise we can follow the proof of Theorem 5.11 and see that the additional constant ω only leads to a larger constant C_2.)

Bernoulli's inequality further implies that $(1+x) \leq e^x$ and hence that

$$r_m(\lambda) = 1 - \lambda\varphi_m(\lambda) = (1-\omega\lambda)^m \leq e^{-\omega\lambda m} \leq 1 =: C_r \qquad \text{for all } m \in \mathbb{N}, \lambda \in (0,\kappa].$$

We now consider for fixed $\nu > 0$ and $m \in \mathbb{N}$ the function $h_m(\lambda) := \lambda^{\nu/2} e^{-\omega\lambda m}$ and compute

$$h_m'(\lambda) = \frac{\nu}{2}\lambda^{\nu/2-1}e^{-\omega\lambda m} - \omega m \lambda^{\nu/2}e^{-\omega\lambda m} = \lambda^{\nu/2-1}e^{-\omega\lambda m}\omega m\left(\frac{\nu}{2\omega m} - \lambda\right).$$

The root $\lambda^* = \frac{\nu}{2\omega m}$ of this derivative satisfies $h_m''(\lambda^*) < 0$, and hence

$$\sup_{\lambda\in(0,\kappa]} \lambda^{\nu/2} r_m(\lambda) \leq \sup_{\lambda\in(0,\infty)} h_m(\lambda) = h_m\left(\frac{\nu}{2\omega m}\right) = e^{-\nu/2}\left(\frac{\nu}{2\omega}\right)^{\nu/2} m^{-\nu/2} =: C_\nu \alpha^{\nu/2}.$$

This shows that (5.8) holds for all $\nu > 0$. Landweber regularization thus has infinite qualification, and the claim follows for $\tau > C_r = 1$ from Theorem 5.11. □

We next study the monotonicity of the Landweber iteration.

Theorem 7.4. *Let $m \in \mathbb{N}$. If $Kx_m^\delta - y^\delta \neq 0$, then*

$$\|Kx_{m+1}^\delta - y^\delta\|_Y < \|Kx_m^\delta - y^\delta\|_Y.$$

Proof. The iteration (7.1) implies that

$$\begin{aligned} Kx_{m+1}^\delta - y^\delta &= K\left((\mathrm{Id} - \omega K^*K)x_m^\delta + \omega K^* y^\delta\right) - y^\delta \\ &= (\mathrm{Id} - \omega KK^*)Kx_m^\delta - (\mathrm{Id} + \omega KK^*)y^\delta \\ &= (\mathrm{Id} - \omega KK^*)(Kx_m^\delta - y^\delta) \end{aligned}$$

and hence due to $\omega < \kappa^{-1} = \sigma_1^{-2} \leq \sigma_n^{-2}$ for all $n \in \mathbb{N}$ that

$$\begin{aligned}\|Kx_{m+1}^\delta - y^\delta\|_Y^2 &= \sum_{n\in\mathbb{N}} (1-\omega\sigma_n^2)^2 \left|\left(Kx_m^\delta - y^\delta \,\middle|\, u_n\right)_Y\right|^2 \\ &< \sum_{n\in\mathbb{N}} \left|\left(Kx_m^\delta - y^\delta \,\middle|\, u_n\right)_Y\right|^2 \leq \|Kx_m^\delta - y^\delta\|_Y^2. \qquad \square\end{aligned}$$

The residual therefore always decreases as $m \to \infty$ (even though a least squares solution minimizing the residual will not exist for $y \notin \mathcal{D}(K^\dagger)$). For the error, this can be guaranteed only up to a certain step.

Theorem 7.5. *Let $m \in \mathbb{N}$. If*

$$\|Kx_m^\delta - y^\delta\|_Y > 2\delta,$$

then

$$\|x_{m+1}^\delta - x^\dagger\|_X < \|x_m^\delta - x^\dagger\|_X.$$

Proof. We use the iteration to write with $\xi_m^\delta = y^\delta - Kx_m^\delta$ and $y = Kx^\dagger$

$$\begin{aligned}\|x_{m+1}^\delta - x^\dagger\|_X^2 &= \|x_m^\delta - x^\dagger + \omega K^*(y^\delta - Kx_m^\delta)\|_X^2 \\ &= \|x_m^\delta - x^\dagger\|_X^2 - 2\omega\left(Kx^\dagger - Kx_m^\delta \,\middle|\, \xi_m^\delta\right)_Y + \omega^2\|K^*\xi_m^\delta\|_X^2 \\ &= \|x_m^\delta - x^\dagger\|_X^2 + \omega\left(\xi_m^\delta - 2y + 2Kx_m^\delta \,\middle|\, \xi_m^\delta\right)_Y + \omega\left(\omega\|K^*\xi_m^\delta\|_X^2 - \|\xi_m^\delta\|_Y^2\right).\end{aligned}$$

We now have to show that the last two terms are negative. For the first term, we use the definition of ξ_m^δ and obtain by inserting $\xi_m^\delta = 2\xi_m^\delta - \xi_m^\delta = 2y^\delta - 2Kx_m^\delta - \xi_m^\delta$ that

$$\begin{aligned}\left(\xi_m^\delta - 2y + 2Kx_m^\delta \,\middle|\, \xi_m^\delta\right)_Y &= 2\left(y^\delta - y \,\middle|\, \xi_m^\delta\right)_Y - \|\xi_m^\delta\|_Y^2 \\ &\leq 2\delta\|\xi_m^\delta\|_Y - \|\xi_m^\delta\|_Y^2 \\ &= \left(2\delta - \|Kx_m^\delta - y^\delta\|_Y\right)\|\xi_m^\delta\|_Y < 0\end{aligned}$$

since the term in parentheses is negative by assumption and $\|\xi_m^\delta\|_Y > 2\delta > 0$.

For the second term, we use $\omega < \kappa^{-1}$ and therefore that

$$\omega\|K^*\xi_m^\delta\|_X^2 \leq \omega\|K^*\|_{\mathbb{L}(Y,X)}^2\|\xi_m^\delta\|_Y^2 = \omega\kappa\|\xi_m^\delta\|^2 < \|\xi_m^\delta\|_X^2$$

and hence

$$\omega\left(\omega\|K^*\xi_m^\delta\|_X^2 - \|\xi_m^\delta\|_Y^2\right) < 0. \tag{7.3}$$

Hence both terms are negative, and the claim follows. $\square$

Hence the Landweber iteration reduces the error until the residual norm drops below twice the noise level. (This implies that for the discrepancy principle, τ should always be chosen less than 2 since otherwise the iteration is guaranteed to terminate *too* early.) From this point on, the error will start to increase again for $y^\delta \notin \mathcal{R}(K)$ by Theorem 5.6. This behavior is called *semiconvergence* and is typical for iterative methods when applied to ill-posed problems. The discrepancy principle then prevents that the error increases arbitrarily. (A slight increase is accepted – how much, depends on the choice of $\tau \in (1, 2)$.)

An important question relating to the efficiency of the Landweber method is the number of steps required for the discrepancy principle to terminate the iteration. The following theorem gives an upper bound.

Theorem 7.6. *Let $\tau > 1$ and $y^\delta \in B_\delta(Kx^\dagger)$. Then the discrepancy principle (7.2) terminates the Landweber iteration (7.1) in step*

$$m(\delta, y^\delta) \leq C\delta^{-2} \qquad \textit{for some } C > 0.$$

Proof. We first derive a convergence rate for the residual norm in terms of m. For this, we consider for $n \geq 0$ the iterate x_n produced by the Landweber iteration applied to the exact data $y := Kx^\delta \in \mathcal{R}(K)$ and denote the corresponding residual by $\xi_n := y - Kx_n$. We now proceed similarly to the proof of Theorem 7.5. Using the iteration (7.1) and (7.3) shows that

$$\begin{aligned} \|x^\dagger - x_n\|_X^2 - \|x^\dagger - x_{n+1}\|_X^2 &= \|x^\dagger - x_n\|_X^2 - \|x^\dagger - x_n - \omega K^*\xi_n\|_X^2 \\ &= 2\omega \left(Kx^\dagger - Kx_n \,\middle|\, \xi_n\right)_Y - \omega^2\|K^*\xi_n\|_X^2 \\ &= \omega\left(\|\xi_n\|_Y^2 - \omega\|K^*\xi_n\|_X^2\right) + \omega\|\xi_n\|_Y^2 \\ &> \omega\|\xi_n\|_Y^2. \end{aligned}$$

Summing over all $n = 0, \dots, m-1$ and using the monotonicity of the residual from Theorem 7.4 then yields

$$\begin{aligned} \|x^\dagger - x_0\|_X^2 - \|x^\dagger - x_m\|_X^2 &= \sum_{n=0}^{m-1} \left(\|x^\dagger - x_n\|_X^2 - \|x^\dagger - x_{n+1}\|_X^2\right) \\ &> \omega \sum_{n=0}^{m-1} \|\xi_n\|_Y^2 > \omega m\|\xi_m\|_X^2. \end{aligned}$$

In particular,

$$\|y - Kx_m\|_Y^2 < (\omega m)^{-1}\|x^\dagger - x_0\|_X^2.$$

As in the proof of Theorem 7.4, we now have due to $x_0 = 0$ that

$$\xi_m^\delta = y^\delta - Kx_m^\delta = (\mathrm{Id} - \omega KK^*)(y^\delta - Kx_{m-1}^\delta) = \dots = (\mathrm{Id} - \omega KK^*)^m y^\delta$$

and similarly for $\xi_m = (\mathrm{Id} - \omega KK^*)^m y$. This yields using $\omega < \kappa^{-1} < \sigma_n^{-2}$ the estimate

$$\|(\mathrm{Id} - \omega KK^*)^m(y^\delta - y)\|_Y^2 = \sum_{n\in\mathbb{N}} (1-\omega\sigma_n^2)^{2m} \left|\left(y^\delta - y \,\middle|\, u_n\right)_Y\right|^2 \leq \|y^\delta - y\|_Y^2$$

and hence that

$$\begin{aligned} \|Kx_m^\delta - y^\delta\|_Y &= \|(\mathrm{Id} - \omega KK^*)^m y^\delta\|_Y \\ &\leq \|(\mathrm{Id} - \omega KK^*)^m y\|_Y + \|(\mathrm{Id} - \omega KK^*)^m (y^\delta - y)\|_Y \\ &\leq \|y - Kx_m\|_Y + \|y^\delta - y\|_Y \\ &\leq (\omega m)^{-1/2}\|x^\dagger - x_0\|_X + \delta. \end{aligned}$$

The discrepancy principle now chooses the stopping index $m(\delta, y^\delta)$ as the first index for which $\|Kx^\delta_{m(\delta,y^\delta)} - y^\delta\|_Y \leq \tau\delta$. Due to the monotonicity of the residual norm, this is the case at the latest for the first $\bar m \in \mathbb{N}$ with

$$(\omega\bar m)^{-1/2}\|x^\dagger - x_0\|_X + \delta \leq \tau\delta;$$

in other words, for which

$$\bar m \geq \omega \frac{\|x^\dagger - x_0\|_X^2}{\omega^2(\tau-1)^2}\delta^{-2} \geq \bar m - 1.$$

This implies that

$$m(\delta, y^\delta) \leq \bar m - 1 \leq C\delta^{-2} + 1$$

with $C := \omega^{-1}(\tau-1)^{-2}\|x^\dagger - x_0\|_X^2 + \max\{1, \delta^2\}$. □

It is not surprising that this estimate can be improved under the usual source condition $x^\dagger \in X_\nu$. Specifically, the estimate (5.15) in the proof of Theorem 5.11 implies for $\alpha = \frac{1}{m}$ the bound $m \leq C\delta^{-\frac{2}{\nu+1}}$. Still, Landweber regularization in practice often requires too many iterations, which motivates accelerated variants such as the one described in [Engl, Hanke & Neubauer 1996, Chapter 6.2, 6.3]. Furthermore, regularization by early stopping can be applied to other iterative methods for solving the normal equation; a particularly popular choice is the conjugate gradient (CG) method; see [Engl, Hanke & Neubauer 1996, Chapter 7].

8 DISCRETIZATION AS REGULARIZATION

And now for something completely different. We have seen that the fundamental difficulty in inverse problems is due to the unboundedness of the pseudoinverse for compact operators $K : X \to Y$ with infinite-dimensional range. It thus suggests itself to construct a sequence $\{K_n\}_{n\in\mathbb{N}}$ of operators with *finite-dimensional* ranges and approximate the wanted minimum norm solution $K^\dagger y$ using the (now continuous) pseudoinverses $(K_n)^\dagger$. This is indeed possible – up to a point. Such finite-dimensional operators can be constructed by either of the following approaches:

(i) We restrict the domain of K to a finite-dimensional subspace $X_n \subset X$ and define $K_n : X_n \to Y$, which has finite-dimensional range because if $\{x_1, \dots, x_n\}$ is a basis of X_n, then $\mathcal{R}(K_n) = \operatorname{span}\{Kx_1, \dots, Kx_n\}$. This approach is referred to as *least-squares projection.*

(ii) We directly restrict the range of K to a finite-dimensional subspace $Y_n \subset Y$ and define $K_n : X \to Y_n$. This approach is referred to as *dual least-squares projection.*

(Of course, we could also restrict domain *and* range and define $K_n : X_n \to Y_n$, but this will not add anything useful from the point of regularization theory.) In this chapter, we will study both approaches, where the second will be seen to have advantages. Since we do not require any spectral theory for this, we will consider again an arbitrary bounded operator $T \in \mathbb{L}(X, Y)$.

8.1 LEAST-SQUARES PROJECTION

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of nested subspaces, i.e.,

$$X_1 \subset X_2 \subset \cdots \subset X,$$

with $\dim X_n = n$ and $\overline{\bigcup_{n\in\mathbb{N}} X_n} = X$. Furthermore, let $P_n := P_{X_n}$ denote the orthogonal projection onto X_n and set $T_n := TP_n \in \mathbb{L}(X, Y)$. Since T_n has finite-dimensional range, $T_n^\dagger := (T_n)^\dagger$ is continuous. We thus define for $y \in Y$ the regularization $x_n := T_n^\dagger y$, i.e., the minimum norm solution of $TP_nx = y$. By Lemma 3.4, we then have

$$x_n \in \mathcal{R}(T_n^\dagger) = \mathcal{N}(T_n)^\perp = \overline{\mathcal{R}(T_n^*)} = \overline{\mathcal{R}(P_nT^*)} \subset X_n$$

since X_n is finite-dimensional and therefore closed and P_n is selfadjoint. (We are thus only looking for a minimum norm solution in X_n instead of in all of X.) To show that $T_n^\dagger$ is a regularization in the sense of Definition 4.1, we have to show that $y \in \mathcal{D}(T^\dagger)$ implies that $T_n^\dagger y \to T^\dagger y$ as $n \to \infty$. This requires an additional assumption.[1]

Lemma 8.1. *Let $y \in \mathcal{D}(T^\dagger)$. Then $x_n \to x^\dagger$ if and only if* $\limsup_{n\to\infty} \|x_n\|_X \leq \|x^\dagger\|_X$.

Proof. If $x_n \to x^\dagger$, the triangle inequality directly yields that

$$\|x_n\|_X \leq \|x_n - x^\dagger\|_X + \|x^\dagger\|_X \to \|x^\dagger\|_X.$$

Conversely, if the lim sup assumption holds, the sequence $\{\|x_n\|_X\}_{n\in\mathbb{N}}$ and thus also $\{x_n\}_{n\in\mathbb{N}}$ is bounded in X. Hence there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ and a $\bar{x} \in X$ with $x_k := x_{n_k} \rightharpoonup \bar{x}$ and $Tx_k \rightharpoonup T\bar{x}$. By the definition of x_k as a least squares solution of $T_k x = y$ (of minimal norm) and by $Tx^\dagger = P_{\overline{\mathcal{R}}} y$ due to Lemma 3.4 (iv) and $\overline{\mathcal{R}}(T) = \mathcal{N}(T)^\perp$, we now have

$$\begin{aligned}\|T_k x_k - Tx^\dagger\|_Y^2 + \|(\mathrm{Id} - P_{\overline{\mathcal{R}}})y\|_Y^2 &= \|T_k x_k - y\|_Y^2 \leq \|T_k x - y\|_Y^2 \\ &= \|T_k x - Tx^\dagger\|_Y^2 + \|(\mathrm{Id} - P_{\overline{\mathcal{R}}})y\|_Y^2 \qquad \text{for all } x \in X.\end{aligned}$$

Since $x_k \in X_k$, we have $x_k = P_k x_k$, and thus $x = P_k x^\dagger$ satisfies

$$\begin{aligned}\|Tx_k - Tx^\dagger\|_Y = \|T_k x_k - Tx^\dagger\|_Y &\leq \|T_k P_k x^\dagger - Tx^\dagger\|_Y = \|TP_k x^\dagger - Tx^\dagger\|_Y \\ &\leq \|T\|_{\mathbb{L}(X,Y)} \|(I - P_k)x^\dagger\|_X.\end{aligned} \tag{8.1}$$

By the assumptions on $\{X_n\}_{n\in\mathbb{N}}$, the last term converges to zero as $k \to \infty$, and hence we have that $Tx_k \to Tx^\dagger$. This implies that $\bar{x} - x^\dagger \in \mathcal{N}(T)$. Now we always have that $x^\dagger \in \mathcal{N}(T)^\perp$, and hence the weak lower semicontinuity of the norm together with the lim sup assumption yields that

$$\|\bar{x} - x^\dagger\|_X^2 + \|x^\dagger\|_X^2 = \|\bar{x}\|_X^2 \leq \liminf_{k\to\infty} \|x_k\|_X^2 \leq \limsup_{k\to\infty} \|x_k\|_X^2 \leq \|x^\dagger\|_X^2,$$

which implies that $\bar{x} = x^\dagger$. This shows that every weakly convergent subsequence has the same limit $x^\dagger$, and therefore the full sequence has to converge weakly to $x^\dagger$. Finally, the lower semicontinuity Combining the weak lower semicontinuity of the norm with the lim sup assumption finally yields that $\|x_n\|_X \to \|x^\dagger\|_X$ as well, and hence the sequence even converges strongly in the Hilbert space X. ☐

Unfortunately, it is possible to construct examples where $\{\|x_n\|_X\}_{n\in\mathbb{N}}$ is not bounded; see, e.g., [Engl, Hanke & Neubauer 1996, Example 3.19]. A sufficient condition for convergence is given in the following theorem.

[1] Here we follow [Kindermann 2016]; the proof in [Engl, Hanke & Neubauer 1996] using a similar equivalence for weak convergence requires an additional assumption, as was pointed out in [Du 2008].

Theorem 8.2. *Let $y \in \mathcal{D}(T^\dagger)$. If*

$$\limsup_{n\to\infty} \|(T_n^*)^\dagger x_n\|_Y = \limsup_{n\to\infty} \|(T_n^\dagger)^* x_n\|_Y < \infty, \tag{8.2}$$

then $x_n \to x^\dagger$ as $n \to \infty$.

Proof. Since

$$\|x_n\|_X^2 = \left(x_n - x^\dagger \,\middle|\, x_n\right)_X + \left(x^\dagger \,\middle|\, x_n\right)_X \le \left(x_n - x^\dagger \,\middle|\, x_n\right)_X + \|x^\dagger\|_X \|x_n\|_X,$$

it suffices to show that the first term on the right-hand side tends to zero as $n \to \infty$. For this, we set $w_n := (T_n^\dagger)^* x_n$ and use that $T_n^* w_n = x_n$ since $\mathcal{R}(T_n^*) \subset X_n$ and therefore $x_n \in \mathcal{R}(T_n^\dagger) = \mathcal{R}(T_n^*)$. This allows us to estimate

$$\begin{aligned} \left(x_n - x^\dagger \,\middle|\, x_n\right)_X &= \left(x_n - x^\dagger \,\middle|\, T_n^* w_n\right)_X = \left(T_n x_n - T_n x^\dagger \,\middle|\, w_n\right)_Y \\ &= \left(T_n x_n - T x^\dagger \,\middle|\, w_n\right)_Y + \left(T x^\dagger - T_n x^\dagger \,\middle|\, w_n\right)_Y \\ &\le \left(\|T_n x_n - T x^\dagger\|_Y + \|T(\mathrm{Id} - P_n) x^\dagger\|_Y\right) \|w_n\|_Y \\ &\le 2\|T\|_{\mathbb{L}(X,Y)} \|(\mathrm{Id} - P_n) x^\dagger\|_X \|w_n\|_Y, \end{aligned} \tag{8.3}$$

where in the last step we have again used (8.1). The last term is now bounded by the assumption (8.2), while the second term and thus the whole right-hand side tend to zero. We can therefore apply Lemma 8.1 to obtain the claim. □

This shows that the least-squares projection only defines a convergent regularization if the subspaces X_n are chosen appropriately for the operator T. Before moving on to the dual least-squares projection (which does not require such a condition), we consider the special case of compact operators.

Theorem 8.3. *If $K \in \mathcal{K}(X, Y)$ and $x^\dagger \in X$ satisfy the condition (8.2), then $x^\dagger \in \mathcal{R}(K^*)$.*

Proof. Setting again $w_n := (K_n^\dagger)^* x_n$, the condition (8.2) implies that $\{w_n\}_{n\in\mathbb{N}}$ is bounded and therefore contains a weakly convergent subsequence $w_k \rightharpoonup \bar{w} \in Y$. Since K and therefore also K^* is compact, $K^* w_k \to K^* \bar{w}$. On the other hand, it follows from $(K_n^\dagger)^* = (K_n^*)^\dagger = (P_n K^*)^\dagger$ that

$$K^* w_k = P_k K^* w_k + (\mathrm{Id} - P_k) K^* w_k = x_k + (\mathrm{Id} - P_k) K^* w_k.$$

Passing to the limit on both sides of the equation and appealing to Theorem 8.2, the boundedness of w_k, and $\|\mathrm{Id} - P_k\|_{\mathbb{L}(X,X)} \to 0$, we deduce that $K^* \bar{w} = x^\dagger$, i.e., $x^\dagger \in \mathcal{R}(K^*)$. □

Hence the condition (8.2) already implies a source condition. It is therefore not surprising that we can give an estimate for the convergence $x_n \to x^\dagger$.

Theorem 8.4. *If $K \in \mathcal{K}(X,Y)$ and $x^\dagger \in X$ satisfy the condition (8.2) and $y \in \mathcal{D}(K^\dagger)$, then there exists a constant $C > 0$ such that*

$$\|x_n - x^\dagger\|_X \leq C\|(\mathrm{Id} - P_n)K^*\|_{\mathbb{L}(Y,X)} \qquad \text{for all } n \in \mathbb{N}.$$

Proof. By Theorem 8.3 there exists a $w \in Y$ with $x^\dagger = K^* w$. Hence, (8.1) implies that

$$\left(x_n - x^\dagger \,\middle|\, x^\dagger\right)_X \leq \|Kx_n - Kx^\dagger\|_Y \|w\|_Y \leq \|K(P_n - \mathrm{Id})x^\dagger\|_Y \|w\|_Y.$$

It follows from this together with (8.3) and the boundedness of the $w_n := (K_n^\dagger)^* x_n$ that

$$\begin{aligned}
\|x_n - x^\dagger\|_X^2 &= \left(x_n - x^\dagger \,\middle|\, x_n\right)_X - \left(x_n - x^\dagger \,\middle|\, x^\dagger\right)_X \\
&\leq 2\|K(\mathrm{Id} - P_n)x^\dagger\|_Y \|w_n\|_Y + \|K(\mathrm{Id} - P_n)x^\dagger\|_Y \|w\|_Y \\
&\leq C\|K(\mathrm{Id} - P_n)x^\dagger\|_Y = C\|K(\mathrm{Id} - P_n)(\mathrm{Id} - P_n)K^* w\|_Y \\
&\leq C\|(\mathrm{Id} - P_n)K^*\|_{\mathbb{L}(Y,X)}^2 \|w\|_Y,
\end{aligned}$$

where we have used in the last step that orthogonal projections are selfadjoint and thus that $(K(\mathrm{Id} - P_n))^* = (\mathrm{Id} - P_n)K^*$. □

8.2 DUAL LEAST-SQUARES PROJECTION

Here we directly discretize the range of T. We thus consider a sequence $\{Y_n\}_{n\in\mathbb{N}}$ of nested subspaces, i.e.,

$$Y_1 \subset Y_2 \subset \cdots \subset \overline{\mathcal{R}(T)} = \mathcal{N}(T^*)^\perp \subset Y,$$

with $\dim Y_n = n$ and $\overline{\bigcup_{n\in\mathbb{N}} Y_n} = \mathcal{N}(T^*)^\perp$. Let now $Q_n := P_{Y_n}$ denote the orthogonal projection onto Y_n and set $T_n := Q_n T \in \mathbb{L}(X, Y_n)$. Again, $T_n^\dagger$ and hence also $T_n^\dagger Q_n$ are continuous, and we can take $x_n := T_n^\dagger Q_n y$ – i.e., the minimum norm solution of $Q_n T x = Q_n y$ – as a candidate for our regularization. To show that this indeed defines a regularization, we introduce the orthogonal projection $P_n := P_{X_n}$ onto

$$X_n := T^* Y_n := \{T^* y \mid y \in Y_n\}.$$

We then have the following useful characterization.

Lemma 8.5. *Let $y \in \mathcal{D}(T^\dagger)$. Then $x_n = P_n x^\dagger$.*

Proof. We first note that by definition of the pseudoinverse and of X_n, we have that

$$\mathcal{R}(T_n^\dagger) = \mathcal{N}(T_n)^\perp = \mathcal{R}(T_n^*) = \mathcal{R}(T^* Q_n) = T^* Y_n = X_n$$

(where the second equation follows from the fact that $\mathcal{R}(T_n^*) = X_n$ is finite-dimensional) and hence that $x_n \in X_n$ as well as $X_n^\perp = \mathcal{N}(T_n)$. This also implies that

$$T_n(\mathrm{Id} - P_n)x = 0 \qquad \text{for all } x \in X,$$

i.e., that $T_n P_n = T_n$. Furthermore, it follows from the fact that $Y_n \subset \mathcal{N}(T^*)^\perp = \overline{\mathcal{R}(T)}$ (and hence that $\mathcal{R}(T)^\perp \subset \mathcal{N}(Q_n)$) together with Lemma 3.4 (iv) that

$$Q_n y = Q_n P_{\overline{\mathcal{R}(T)}} y = Q_n T T^\dagger y = Q_n T x^\dagger = T_n x^\dagger.$$

We thus obtain for any $x \in X$ that

$$\|T_n x - Q_n y\|_Y = \|T_n x - T_n x^\dagger\|_Y = \|T_n x - T_n P_n x^\dagger\|_Y = \|T_n(x - P_n x^\dagger)\|_Y.$$

Now x_n is defined as the minimum norm solution of $T_n x = Q_n y$, i.e., as the one $x \in \mathcal{N}(T_n)^\perp = X_n$ minimizing $\|T_n x - Q_n y\|_Y$ – which is obviously minimal for $x = P_n x^\dagger \in X_n$. Since the minimum norm solution is unique, we have that $x_n = P_n x^\dagger$. □

Theorem 8.6. *Let $y \in \mathcal{D}(T^\dagger)$. Then $x_n \to x^\dagger$.*

Proof. The construction of Y_n implies that $X_n \subset X_{n+1}$ and hence that

$$\overline{\bigcup_{n\in\mathbb{N}} X_n} = \overline{\bigcup_{n\in\mathbb{N}} T^* Y_n} = \overline{T^* \bigcup_{n\in\mathbb{N}} Y_n} = \overline{T^* \mathcal{N}(T^*)^\perp} = \overline{\mathcal{R}(T^*)} = \mathcal{N}(T)^\perp.$$

Using $x^\dagger \in \mathcal{R}(T^\dagger) = \mathcal{N}(T)^\perp$, we deduce that $x_n \to x^\dagger$. □

Under a source condition, we can show a similar error estimate as in Theorem 8.4.

Theorem 8.7. *Let $T \in \mathbb{L}(X, Y)$ and $y \in \mathcal{D}(T^\dagger)$. If $x^\dagger = T^\dagger y \in \mathcal{R}(T^*)$, then there exists a constant $C > 0$ such that*

$$\|x_n - x^\dagger\|_X \le C\|(\mathrm{Id} - P_n)T^*\|_{\mathbb{L}(Y,X)} \qquad \textit{for all } n \in \mathbb{N}.$$

Proof. The source condition $x^\dagger = T^* w$ for some $w \in Y$ and Lemma 8.5 immediately yield that

$$\|x_n - x^\dagger\|_X = \|P_n x^\dagger - x^\dagger\|_X = \|(\mathrm{Id} - P_n)T^* w\|_X \le \|(\mathrm{Id} - P_n)T^*\|_{\mathbb{L}(Y,X)} \|w\|_Y. \qquad \square$$

The dual least-squares projection thus defines a regularization operator as well. By Theorem 4.5, there thus exists (at least for compact operators) an a priori choice rule that turns the dual least-squares projection into a convergence regularization method. Characterizing this choice rule requires estimating the norm of $T_n^\dagger$, for which we can use that T_n has finite-dimensional range and is therefore compact. Hence there exists a (finite) singular system $\{(\mu_k, \tilde{u}_k, \tilde{v}_k\}_{k\in\{1,\dots,n\}}$; in particular, we can use that μ_n is the smallest (by magnitude) singular value of T_n.

Theorem 8.8. *Let $y \in \mathcal{D}(T^\dagger)$ and for $y^\delta \in B_\delta(y)$ set $x_n^\delta := T_n^\dagger Q_n y$. If $n(\delta)$ is chosen such that*

$$n(\delta) \to \infty, \qquad \frac{\delta}{\mu_{n(\delta)}} \to 0 \qquad \text{for } \delta \to 0,$$

then $x^\delta_{n(\delta)} \to x^\dagger$ as $\delta \to 0$.

Proof. We proceed as in the proof of Theorem 4.6 and use the standard error decomposition

$$\|x^\delta_{n(\delta)} - x^\dagger\|_X \leq \|x_{n(\delta)} - x^\dagger\|_X + \|x^\delta_{n(\delta)} - x_{n(\delta)}\|_X.$$

By Theorem 8.7, the first term tends to zero as $n \to \infty$.

For the second term, we use the singular value decomposition of T_n and (3.14) to obtain for any $n \in \mathbb{N}$ that

$$\|T_n^\dagger y\|_X^2 = \sum_{k=1}^n \mu_k^{-2} |\,(y \mid \tilde u_k)_Y\,|^2 \leq \mu_n^{-2}\|y\|_Y^2 \qquad \text{for all } y \in Y,$$

with equality for $y = \tilde u_n \in Y$. This implies that $\|T_n^\dagger\|_{\mathbb{L}(Y,X)} = \mu_n^{-1}$. Since Q_n is an orthogonal projection, we have that

$$\|x_n^\delta - x_n\|_X = \|T_n^\dagger Q_n(y^\delta - y)\|_X \leq \|T_n^\dagger\|_{\mathbb{L}(Y,X)}\|y^\delta - y\|_Y \leq \frac{\delta}{\mu_n}.$$

The claim now follows from the assumptions on $n(\delta)$. □

Under the source condition from Theorem 8.7, we can in this way also obtain convergence rates as in Theorem 5.9. (Similar results also hold for the least-squares projection under the additional assumption (8.2).)

We can now ask how to choose Y_n for given $n \in \mathbb{N}$ in order to minimize the regularization error, which by Theorem 8.8 entails minimizing μ_n. This question can be answered explicitly for compact operators.

Theorem 8.9. *Let $K \in \mathcal{K}(X,Y)$ have the singular system $\{(\sigma_n, u_n, v_n)\}_{n\in\mathbb{N}}$. If $Y_n \subset Y$ with* $\dim Y_n = n$*, then $\mu_n \leq \sigma_n$.*

Proof. If μ_n is a singular value of K_n, then μ_n^2 is an eigenvalue of $K_nK_n^* = Q_nKK^*Q_n$; similarly, σ_n^2 is an eigenvalue of KK^*. Set now $U_k := \operatorname{span}\{u_1,\dots,u_k\} \subset \overline{\mathcal{R}(K^*)}$ for all $k \in \mathbb{N}$. Since $\dim Y_n = n$, there exists $\bar y \in U_{n-1}^\perp \cap Y_n$ with $\|\bar y\|_Y = 1$ (otherwise $U_{n-1}^\perp \subset Y_n^\perp$, but this is impossible since the codimension U_{n-1} is too small). The Courant–Fischer min–max principle (2.4) thus implies that

$$
\begin{aligned}
\mu_n^2 &= \max_V \min_y \left\{ (Q_n K K^* Q_n y \mid y)_y \;\middle|\; \|y\|_Y = 1,\ y \in V,\ \dim V = n \right\} \\
&= \min_y \left\{ (KK^* y \mid y)_Y \;\middle|\; \|y\|_Y = 1,\ y \in Y_n \right\} \le (KK^* \bar{y} \mid \bar{y})_Y \\
&\le \max_y \left\{ (KK^* y \mid y)_Y \;\middle|\; \|y\|_Y = 1,\ y \in U_{n-1}^\perp \right\} = \sigma_n^2
\end{aligned}
$$

since the maximum is attained for $y = u_n \in U_{n-1}^\perp$. □

The proof also shows that equality of the singular values holds for $Y_n = U_n$, because then $\bar{y} = u_n$ is the only vector that is a candidate for minimization or maximization. But this choice corresponds exactly to the truncated singular value decomposition from Example 5.2 (i). In fact, the choice $Y_n = U_n$ is optimal with respect to the approximation error as well.

Theorem 8.10. *Let $K \in \mathcal{K}(X, Y)$ have the singular system $\{(\sigma_n, u_n, v_n)\}_{n\in\mathbb{N}}$. If $Y_n \subset Y$ with $\dim Y_n = n$, then*

$$\|(\mathrm{Id} - P_n)K^*\|_{\mathbb{L}(Y,X)} \ge \sigma_{n+1},$$

with equality for $Y_n = U_n$.

Proof. We again use the Courant–Fischer min–max principle, this time for the eigenvalue σ_n^2 of K^*K. Setting $X_n := K^* Y_n$ and $P_n := P_{X_n}$, we then have that

$$
\begin{aligned}
\sigma_{n+1}^2 &= \min_V \max_x \{ (K^*Kx \mid x)_X \mid \|x\|_X = 1 \}\ x \in V \subset X,\ \dim V^\perp = n \\
&\le \max_x \{ (K^*Kx \mid x)_X \mid \|x\|_X = 1 \}\ x \in X_n^\perp \\
&= \max_x \{ (K^*K(I - P_n)x \mid (I - P_n)x)_X \mid \|x\|_X = 1 \} \\
&= \max_x \left\{ \|K(I - P_n)x\|_Y^2 \;\middle|\; \|x\|_X = 1 \right\} \\
&= \|K(\mathrm{Id} - P_n)\|_{\mathbb{L}(X,Y)}^2 = \|(\mathrm{Id} - P_n)K^*\|_{\mathbb{L}(Y,X)}^2.
\end{aligned}
$$

If $Y_n = U_n$, then $X_n = K^* U_n = \operatorname{span}\{v_1, \dots, v_n\}$, and the minimum in the inequality is attained for this subspace. □

Hence the best possible convergence rate (under the source condition from Theorem 8.7) for the dual least-squares projection is

$$\|x_n^\delta - x^\dagger\|_X \le C \left(\sigma_{n+1} + \frac{\delta}{\sigma_n} \right),$$

and this rate is attained for the truncated singular value decomposition.

Without knowledge of a singular system, however, it is necessary in practice to choose n very small in order to ensure the condition on μ_n. But this leads to a very coarse discretization that does not sufficiently capture the behavior of the infinite-dimensional operator. The usual approach is therefore to combine a much finer discretization with one of the regularization methods discussed in the previous chapters. To obtain an optimal convergence rate and to avoid needless computational effort, one should then appropriately choose the regularization parameter in dependence of δ as well as of n (or, vice versa, choose n in dependence of $\alpha(\delta)$).

Part III

NONLINEAR INVERSE PROBLEMS

9 NONLINEAR ILL-POSED PROBLEMS

We now consider *nonlinear* operators $F : U \to Y$ for $U \subset X$ and Hilbert spaces X and Y. The corresponding nonlinear inverse problem then consists in solving the operator equation $F(x) = y$. Such problems occur in many areas; in particular, trying to reconstruct the coefficients of a partial differential equations from a solution for given data (right-hand sides, initial or boundary conditions), e.g., in electrical impedance tomography, is a nonlinear ill-posed problem. Here we will characterize this ill-posedness in an abstract setting; concrete examples would require results on partial differential equations that would go far beyond the scope of these notes.

A fundamental difference between linear and nonlinear operators is that the latter can act very differently on different subsets of X. The global characterization of well- or ill-posedness in the sense of Hadamard is hence too restrictive. We therefore call the operator $F : U \to y$ *locally well-posed* in $x \in U$ if there exists an $r > 0$ such that for all sequences $\{x_n\}_{n\in\mathbb{N}} \subset B_r(x) \cap U$ with $F(x_n) \to F(x)$, we also have that $x_n \to x$. Otherwise the operator is called *locally ill-posed* (in x). In this case, there exists for all $r > 0$ a sequence $\{x_n\}_{n\in\mathbb{N}} \subset B_r(x) \cap U$ with $F(x_n) \to F(x)$ such that x_n does not converge to x. A linear operator $T : X \to Y$ is either locally well-posed for all $x \in X$ or locally ill-posed for all $x \in X$. The latter holds if and only if T is not injective or $\mathcal{R}(T)$ is not closed (e.g., for compact operators with infinite-dimensional range).[1] For nonlinear operators, the situation is a bit more involved. As in the linear case, we call $F : U \to Y$ compact, if every bounded sequence $\{x_n\}_{n\in\mathbb{N}} \subset U$ admits a convergent subsequence of $\{F(x_n)\}_{n\in\mathbb{N}} \subset Y$. However, nonlinear compact operators need *not* be continuous and hence completely continuous (consider, e.g., an arbitrary bounded operator with finite-dimensional range); the latter is therefore an additional assumption. In fact, a weaker assumption suffices: an operator $F : U \to X$ is called *weakly closed*, if $x_n \rightharpoonup x \in U$ and $F(x_n) \rightharpoonup y$ imply that $F(x) = y$.

Lemma 9.1. *Let $F : U \to Y$ be compact and weakly closed. Then F is completely continuous, i.e., maps weakly convergent sequences in X to strongly convergent sequences in Y.*

Proof. Let $\{x_n\}_{n\in\mathbb{N}} \subset U$ be a weakly converging sequence with $x_n \rightharpoonup x \in U$. Then $\{x_n\}_{n\in\mathbb{N}}$ is bounded, and hence $\{F(x_n)\}_{n\in\mathbb{N}}$ contains a convergent subsequence $\{F(x_{n_k})\}_{k\in\mathbb{N}}$ with $F(x_{n_k}) \to y \in Y$. Since strongly convergent sequences also converge weakly (to the same

[1] The local ill-posedness thus generalizes the (global) ill-posedness in the sense of Nashed, not of Hadamard.

limit), the weak closedness yields that $y = F(x)$. Hence the limit is independent of the subsequence, which implies that the whole sequence converges. □

For such operators, we can show an analogous result to Corollary 3.8.

Theorem 9.2. *Let X be an infinite-dimensional separable Hilbert space and $U \subset X$. If $F : U \to Y$ is completely continuous, then F is locally ill-posed in all interior points of U.*

Proof. Since X is separable, there exists an (infinite) orthonormal basis $\{u_n\}_{n\in\mathbb{N}}$. Let now $x \in U$ be an interior point and define for $r > 0$ with $B_r(x) \subset U$ the points $x_n := x + \frac{r}{2}u_n \in B_r(x)$. Then $\|x_n - x\|_X = \frac{r}{2}$, but the fact that $u_n \rightharpoonup 0$ for any orthonormal basis implies that $x_n \rightharpoonup x$ and hence that $F(x_n) \to F(x)$ due to the complete continuity of F. □

As in the linear case we now define minimum norm solutions and regularizations. Since $0 \in X$ can now longer be taken as a generic point, we denote for given $y \in \mathcal{R}(F)$ and $x_0 \in X$ any point $x^\dagger \in U$ with $F(x^\dagger) = y$ and

$$\|x^\dagger - x_0\|_X = \min\{\|x - x_0\|_X \mid F(x) = y\}$$

as x_0*-minimum norm solution*. For nonlinear inverse problems, these need not be unique in contrast to the linear case. Their existence also requires that $F(x) = y$ actually admits a solution. A regularization of $F(x) = y$ is now a family $\{R_\alpha\}_{\alpha>0}$ of continuous (possibly nonlinear) operators $R_\alpha : X \times Y \to X$ such that $R_\alpha(x_0, y)$ converges to an x_0-minimum norm solution as $\alpha \to 0$. In combination with a parameter choice rule for α, we define (convergent) regularization methods as before. For nonlinear inverse problems, these operators can in general not be given explicitly; most regularizations are instead based on an (iterative) linearization of the problem.

This requires a suitable notion of derivatives for operators between normed vector spaces. Let X, Y be normed vector spaces, $F : U \to Y$ be an operator with $U \subset X$ and $x \in U$, and $h \in X$ be arbitrary.

- If the one-sided limit

  $$F'(x; h) := \lim_{t\to 0^+} \frac{F(x + th) - F(x)}{t} \in Y,$$

  exists, it is called the *directional derivative* of F in x in direction h.

- If $F'(x; h)$ exists for all $h \in X$ and

  $$DF(x) : X \to Y, h \mapsto F'(x; h)$$

  defines a bounded linear operator, we call F *Gâteaux differentiable* (in x) and $DF \in \mathbb{L}(X, Y)$ its *Gâteaux derivative*.

- If additionally
$$\lim_{\|h\|_X\to 0} \frac{\|F(x+h)-F(x)-DF(x)h\|_Y}{\|h\|_X} = 0,$$
then F is called *Fréchet differentiable* (in x) and $F'(x) := DF(x) \in \mathbb{L}(X,Y)$ its *Fréchet derivative*.

- If the mapping $F' : U \to \mathbb{L}(X,Y)$, $x \mapsto F'(x)$, is (Lipschitz) continuous, we call F *(Lipschitz) continuously differentiable*.

The difference between Gâteaux and Fréchet differentiable lies in the approximation error of F near x by $F(x) + DF(x)h$: While it only has to be bounded in $\|h\|_X$ – i.e., linear in $\|h\|_X$ – for a Gâteaux differentiable function, it has to be superlinear in $\|h\|_X$ if F is Fréchet differentiable. (For a *fixed* direction h, this of course also the case for Gâteaux differentiable functions; Fréchet differentiability thus additionally requires a uniformity in h.)

If F is Gâteaux differentiable, the Gâteaux derivative can be computed via
$$DF(x)h = \left(\tfrac{d}{dt}F(x+th)\right)\Big|_{t=0}.$$
(However, the existence and linearity of this limit does *not* show the Gâteaux differentiability of F since it doesn't imply that $DF(x)$ is continuous with respect to the right norms.) Bounded linear operators $F \in \mathbb{L}(X,Y)$ are obviously Fréchet differentiable with derivative $F'(x) = F \in \mathbb{L}(X,Y)$ for all $x \in X$. Note that the Gâteaux derivative of a functional $F : X \to \mathbb{R}$ is an element of the *dual space* $X^* = \mathbb{L}(X,\mathbb{R})$ and thus cannot be added to elements in X. However, in Hilbert spaces (and in particular in $\mathbb{R}^n$), we can use the Fréchet–Riesz Theorem 2.2 to identify $DF(x) \in X^*$ with an element $\nabla F(x) \in X$, called *gradient* of F, in a canonical way via
$$DF(x)h = (\nabla F(x) \mid h)_X \qquad \text{for all } h \in X.$$
As an example, let us consider the functional $F(x) = \frac{1}{2}\|x\|_X^2$, where the norm is induced by the inner product. Then we have for all $x, h \in X$ that
$$F'(x;h) = \lim_{t\to 0^+} \frac{\frac{1}{2}(x+th \mid x+th)_X - \frac{1}{2}(x \mid x)_X}{t} = (x \mid h)_X = DF(x)h,$$
since the inner product is linear in h for fixed x. Hence, the squared norm is Gâteaux differentiable in x with derivative $DF(x) = h \mapsto (x \mid h)_X \in X^*$ and gradient $\nabla F(x) = x \in X$; it is even Fréchet differentiable since
$$\lim_{\|h\|_X\to 0} \frac{\left|\frac{1}{2}\|x+h\|_X^2 - \frac{1}{2}\|x\|_X^2 - (x \mid h)_X\right|}{\|h\|_X} = \lim_{\|h\|_X\to 0} \frac{1}{2}\|h\|_X = 0.$$
If the same mapping is now considered on a smaller Hilbert space $X' \hookrightarrow X$ (e.g., $X = L^2(\Omega)$ and $X' = H^1(\Omega)$), then the derivative $DF(x) \in (X')^*$ is still given by $DF(x)h = (x \mid h)_X$ (now only for all $h \in X'$), but the gradient $\nabla F \in X'$ is now characterized by
$$DF(x)h = (\nabla F(x) \mid h)_{X'} \qquad \text{for all } h \in X'.$$

Different inner products thus lead to different gradients.

Further derivatives can be obtained through the usual calculus, whose proof in Banach spaces is exactly as in $\mathbb{R}^n$. As an example, we prove a chain rule.

Theorem 9.3. *Let X, Y, and Z be Banach spaces, and let $F : X \to Y$ be Fréchet differentiable in $x \in X$ and $G : Y \to Z$ be Fréchet differentiable in $y := F(x) \in Y$. Then, $G \circ F$ is Fréchet differentiable in x and*

$$(G \circ F)'(x) = G'(F(x)) \circ F'(x).$$

Proof. For $h \in X$ with $x + h \in \operatorname{dom} F$ we have

$$(G \circ F)(x + h) - (G \circ F)(x) = G(F(x + h)) - G(F(x)) = G(y + g) - G(y)$$

with $g := F(x + h) - F(x)$. The Fréchet differentiability of G thus implies that

$$\|(G \circ F)(x + h) - (G \circ F)(x) - G'(y)g\|_Z = r_1(\|g\|_Y)$$

with $r_1(t)/t \to 0$ for $t \to 0$. The Fréchet differentiability of F further implies

$$\|g - F'(x)h\|_Y = r_2(\|h\|_X)$$

with $r_2(t)/t \to 0$ for $t \to 0$. In particular,

$$\|g\|_Y \le \|F'(x)h\|_Y + r_2(\|h\|_X). \tag{9.1}$$

Hence, with $c := \|G'(F(x))\|_{\mathbb{L}(Y,Z)}$ we have

$$\|(G \circ F)(x + h) - (G \circ F)(x) - G'(F(x))F'(x)h\|_Z \le r_1(\|g\|_Y) + c\, r_2(\|h\|_X).$$

If $\|h\|_X \to 0$, we obtain from (9.1) and $F'(x) \in \mathbb{L}(X, Y)$ that $\|g\|_Y \to 0$ as well, and the claim follows. □

A similar rule for Gâteaux derivatives does not hold, however.

We will also need the following variant of the mean value theorem. Let $[a, b] \subset \mathbb{R}$ be a bounded interval and $f : [a, b] \to X$ be continuous. We then define the *Bochner integral* $\int_a^b f(t)\,dt \in X$ using the Fréchet–Riesz Theorem 2.2 via

$$\left(\int_a^b f(t)\,dt \,\middle|\, z\right)_X = \int_a^b (f(t) \mid z)_X \,dt \qquad \text{for all } z \in X, \tag{9.2}$$

since by the continuity of $t \mapsto \|f(t)\|_X$ on the compact interval $[a, b]$, the right-hand side defines a continuous linear functional on X. The construction then directly implies that

$$\left\|\int_a^b f(t)\,dt\right\|_X \le \int_a^b \|f(t)\|_X\,dt. \tag{9.3}$$

Theorem 9.4. *Let $F : U \to Y$ be Fréchet differentiable, and let $x \in U$ and $h \in Y$ be given with $x + th \in U$ for all $t \in [0, 1]$. Then*

$$F(x+h) - F(x) = \int_0^1 F'(x+th)h\,dt.$$

Proof. Consider for arbitrary $y \in Y$ the function

$$f : [0,1] \to \mathbb{R}, \qquad t \mapsto (F(x+th) \mid y)_Y\,.$$

From Theorem 9.3 we obtain that f (as a composition of operators between normed vector spaces) is differentiable with

$$f'(t) = (F'(x+th)h \mid y)_Y\,,$$

and the fundamental theorem of calculus in $\mathbb{R}$ yields that

$$(F(x+h) - F(x) \mid y)_Y = f(1) - f(0) = \int_0^1 f'(t)\,dt = \left(\int_0^1 F'(x+th)h\,dt \,\middle|\, y\right)_Y,$$

where the last equality follows from (9.2). Since $y \in Y$ was arbitrary, the claim follows. □

If the Fréchet derivative is locally Lipschitz continuous, i.e., if there exist $L > 0$ and $\delta > 0$ such that

$$\|F'(x_1) - F'(x_2)\|_{\mathbb{L}(X,Y)} \le L\|x_1 - x_2\|_X \qquad \text{for all } x_1, x_2 \in B_\delta(x), \tag{9.4}$$

the linearization error can even be estimated quadratically.

Lemma 9.5. *Let $F : U \to Y$ Lipschitz continuously differentiable in a neighborhood $V \subset U$ of $x \in U$. Then for all $h \in X$ with $x + th \in V$ for $t \in [0, 1]$,*

$$\|F(x+h) - F(x) - F'(x)h\|_Y \le \frac{L}{2}\|h\|_X^2.$$

Proof. Theorem 9.4 together with (9.3) and (9.4) directly yield that

$$\begin{aligned}\|F(x+h) - F(x) - F'(x)h\|_Y &= \left\|\int_0^1 F'(x+th)h - F'(x)h\,dt\right\|_Y \\ &\le \int_0^1 \|F'(x+th)h - F'(x)h\|\,dt \\ &\le \int_0^1 Lt\|h\|_X^2\,dt = \frac{L}{2}\|h\|_X^2.\end{aligned}$$

□

A natural question is now about the relationship between the local ill-posedness of $F : U \to Y$ in x and of its linearization $F'(x) \in \mathbb{L}(X, Y)$. The following result suggests that at least for completely continuous operators, the latter inherits the ill-posedness of the former.

Theorem 9.6. *If $F : U \to Y$ is completely continuous and Fréchet differentiable in $x \in U$, then $F'(x) \in \mathbb{L}(X, Y)$ is compact.*

Proof. Let $x \in U$ be arbitrary and assume to the contrary that $F'(x)$ is not compact and therefore not completely continuous. Then there exists a sequence $\{h_n\}_{n\in\mathbb{N}}$ with $h_n \rightharpoonup 0$ as well as an $\varepsilon > 0$ such that

$$\|F'(x)h_n\|_Y \geq \varepsilon \qquad \text{for all } n \in \mathbb{N}.$$

Since weak convergence implies boundedness, we can assume without loss of generality (by proper scaling of h_n and ε) that $\|h_n\|_X \leq 1$ for all $n \in \mathbb{N}$. By definition of the Fréchet derivative, there then exists a $\delta > 0$ such that

$$\|F(x+h) - F(x) - F'(x)h\|_Y \leq \frac{\varepsilon}{2}\|h\|_X \qquad \text{for all } \|h\|_X \leq \delta.$$

Since $\{h_n\}_{n\in\mathbb{N}}$ is bounded, there exists a $\tau > 0$ sufficiently small that $\|\tau h_n\|_X \leq \delta$ and $x + \tau h_n \in U$ for all $n \in \mathbb{N}$ (otherwise F would not be differentiable in x). Then we have that $x + \tau h_n \rightharpoonup x$; however, for all $n \in \mathbb{N}$,

$$\begin{aligned}\|F(x+\tau h_n) - F(x)\|_Y &= \|F'(x)(\tau h_n) + F(x+\tau h_n) - F(x) - F'(x)(\tau h_n)\|_Y \\ &\geq \|F'(x)(\tau h_n)\|_Y - \|F(x+\tau h_n) - F(x) - F'(x)(\tau h_n)\|_Y \\ &\geq \tau\varepsilon - \tau\|h_n\|_X\frac{\varepsilon}{2} \geq \tau\frac{\varepsilon}{2}.\end{aligned}$$

Hence F is not completely continuous. □

Note that this does not necessarily imply that $F'(x)h = y - F(x+h)$ is ill-posed, as $F'(x)$ may happen to have finite-dimensional range. Conversely, a locally well-posed problem may have an ill-posed linearization; see [Engl, Kunisch & Neubauer 1989, Example A.1, A.2]. This naturally has consequences to any regularization that relies on linearization. The reason for this discrepancy is the fact that although the linearization error tends to zero superlinearly as $\|h\|_X \to 0$, for fixed $h \in X$ the error may be much larger than either the nonlinear residual $y - F(x)$ or the linear residual $y - F(x+h) - F'(x)h$. To obtain stronger results, we thus have to impose conditions on the nonlinearity of F.

One possibility is to require more smoothness of F, e.g., local Lipschitz continuity of the derivative around $x \in U$. Under this assumption, the linearization indeed inherits the local ill-posedness.

Theorem 9.7. *Let $F : U \to Y$ be Fréchet differentiable with locally Lipschitz continuous derivative. If F is locally ill-posed in $x \in U$, then $F'(x)$ is locally ill-posed in all $h \in \mathcal{N}(F'(x))$.*

Proof. Assume to the contrary that the nonlinear operator is locally ill-posed but its linearization is locally well-posed. The latter is equivalent to $F'(x)$ being injective and having closed range. Hence by Theorem 3.7 there exists a continuous pseudoinverse $F'(x)^\dagger \in \mathbb{L}(Y,X)$. Now if $F'(x)^\dagger$ is continuous, so is $(F'(x)^*)^\dagger = (F'(x)^\dagger)^*$, and we can thus find for all $h \in X$ a $w := (F'(x)^*)^\dagger h \in Y$ with $\|w\|_Y \leq C\|h\|_X$. Letting $\mu \in (0,1)$ and setting $\delta := \frac{2\mu}{CL}$, we then have in particular that $\|w\|_Y \leq \frac{2\mu}{L}$ for all $\|h\|_X \leq \delta$. Furthermore, Lemma 3.4 (iv) together with $\mathcal{R}(F'(x)^*) = \overline{\mathcal{R}(F'(x)^*)} = \mathcal{N}(F'(x))^\perp = X$ (since if $(F'(x)^*)^\dagger$ is continuous, $F'(x)^*$ has closed range as well) implies that

$$F'(x)^* w = F'(x)^*(F'(x)^*)^\dagger h = h.$$

We now bound the linearization error with the help of this "linearized source condition" and Lemma 9.5: For all $h \in X$ with $\|h\|_X \leq \delta$, we have that

$$\begin{aligned}\|F(x+h) - F(x) - F'(x)h\|_Y &\leq \frac{L}{2}\|h\|_X^2 = \frac{L}{2}\|F'(x)^*w\|_X^2 = \frac{L}{2}\left(F'(x)F'(x)^*w \mid w\right)_Y \\ &\leq \frac{L}{2}\|F'(x)F'(x)^*w\|_Y\|w\|_Y \\ &\leq \mu\|F'(x)h\|_Y.\end{aligned}$$

The triangle inequality then yields that

$$\begin{aligned}\|F'(x)h\|_Y &= \|F(x+h) - F(x) - F'(x)h - F(x+h) + F(x)\|_Y \\ &\leq \mu\|F'(x)h\|_Y + \|F(x+h) - F(x)\|_Y\end{aligned}$$

and hence that

$$\|F'(x)h\|_Y \leq \frac{1}{1-\mu}\|F(x+h) - F(x)\|_Y \qquad \text{for all } \|h\|_X \leq \delta. \tag{9.5}$$

Since we have assumed that F is locally ill-posed, there has to exist a sequence $\{h_n\}_{n\in\mathbb{N}}$ with $\|x + h_n - x\|_X = \|h_n\|_X = \frac{\delta}{2}$ but $F(x + h_n) \to F(x)$. But from (9.5), we then obtain that $F'(x)(x + h_n - x) = F'(x)h_n \to 0$, in contradiction to the assumed local well-posedness of the linearization. □

An alternative to (9.4) is the so-called *tangential cone condition*: For given $x \in U$, there exist $\eta < 1$ and $\delta > 0$ such that

$$\|F(x+h) - F(x) - F'(x)h\|_Y \leq \eta\|F(x+h) - F(x)\|_Y \qquad \text{for all } \|h\|_X \leq \delta. \tag{9.6}$$

In other words, the linearization error should be *uniformly* bounded by the nonlinear residual. Here we can even show equivalence.

Theorem 9.8. *Let $F : U \to Y$ be Fréchet differentiable and satisfy the tangential cone condition (9.6) in $x \in U$. Then Fy is locally ill-posed in $x \in U$ if and only if $F'(x)$ is locally ill-posed (in any $h \in X$).*

Proof. From the tangential cone condition together with the (standard and reverse) triangle inequalities, we obtain that

$$(9.7) \qquad (1-\eta)\|F(x+h) - F(x)\|_Y \le \|F'(x)h\|_Y \le (1+\eta)\|F(x+h) - F(x)\|_Y$$

for all $\|h\|_X \le \delta$. The second inequality coincides with (9.5), which we have already shown to imply the local ill-posedness of the linearization of a locally ill-posed nonlinear operator. We can argue similarly for the first inequality: Assume that $F'(x)$ is locally ill-posed. Then there exists a sequence $\{h_n\}_{n\in\mathbb{N}}$ with $\|x + h_n - x\|_X = \|h_n\| = \frac{\delta}{2}$ but $F'(x)h_n \to 0$, which together with (9.7) implies that $F(x + h_n) \to F(x)$ as well. Hence, F is also ill-posed. □

In combination with a weak source condition, the tangential cone condition even implies local uniqueness of the x_0-minimum norm solution.

Theorem 9.9. *Let $F : U \to Y$ be Fréchet differentiable and $y \in Y$ and $x_0 \in X$ be given. If the tangential cone condition (9.6) holds in $x^\dagger \in U$ with $F(x^\dagger) = y$ and $x^\dagger - x_0 \in \mathcal{N}(F'(x^\dagger))^\perp$, then $x^\dagger$ is the unique x_0-minimum norm solution in $B_\delta(x^\dagger)$ for the $\delta > 0$ from (9.6).*

Proof. Let $x \in B_\delta(x^\dagger) \setminus \{x^\dagger\}$ with $F(x) = y$ be arbitrary. Then (9.6) for $h := x - x^\dagger$ implies that $F'(x^\dagger)(x - x^\dagger) = 0$, i.e., that $x - x^\dagger \in \mathcal{N}(F'(x^\dagger)) \setminus \{0\}$. It follows that

$$\begin{aligned}
\|x - x_0\|_X^2 &= \|x^\dagger - x_0 + x - x^\dagger\|_X^2 \\
&= \|x^\dagger - x_0\|_X^2 + 2\left(x^\dagger - x_0 \,\middle|\, x - x^\dagger\right)_X + \|x - x^\dagger\|_X^2 \\
&> \|x^\dagger - x_0\|_X^2
\end{aligned}$$

since the inner product vanishes due to orthogonality and we have assumed that $x \neq x^\dagger$. Hence $x^\dagger$ is the (locally) unique x_0-minimum norm solution. □

It should be admitted that it is often very difficult to verify these abstract conditions for concrete nonlinear inverse problems; there are even examples where these can be show *not* to hold. Thus one often uses strongly problem-specific approaches instead of an abstract theory for nonlinear problems.[2] Still, the abstract perspective can be useful by showing limits and possibilities.

[2]"Linear inverse problems are all alike; every nonlinear inverse problem is nonlinear in its own way."

10 TIKHONOV REGULARIZATION

The starting point of Tikhonov regularization of nonlinear inverse problems $F(x) = y$ is Theorem 6.5: For given $\alpha > 0$, $x_0 \in X$, and $y \in Y$, we choose x_α as minimizer of the Tikhonov functional

$$J_\alpha(x) := \frac{1}{2}\|F(x) - y\|_Y^2 + \frac{\alpha}{2}\|x - x_0\|_X^2. \tag{10.1}$$

If F is not linear, we cannot express this choice through an explicit regularization operator R_α. We thus have to proceed differently to show existence of a solution, continuous dependence of x_α on y, and convergence to an x_0 minimum norm solution as $\alpha \to 0$. On the other hand, this is possible under weaker assumptions: It suffices to require that F is weakly closed with non-empty and weakly closed domain $\operatorname{dom} F =: U$ (which we always assume from here on). These assumptions also ensure for $y \in \mathcal{R}(F)$ the existence of a (not necessarily unique) x_0-minimum norm solution $x^\dagger \in U$.

We first show existence of a minimizer. The proof is a classical application of Tonelli's *direct method* of the calculus of variations, which generalizes the Weierstraß Theorem (every continuous function attains its minimum and maximum on a finite-dimensional compact set) to infinite-dimensional vector spaces.

Theorem 10.1. *Let $F : U \to Y$ be weakly closed, $\alpha > 0$, $x_0 \in X$, and $y \in Y$. Then there exists a minimizer $x_\alpha \in U$ of J_α.*

Proof. We first note that $J_\alpha(x) \geq 0$ for all $x \in U$. Hence the set $\{J_\alpha(x) \mid x \in U\} \subset \mathbb{R}$ is bounded from below and thus has a finite infimum. This implies that there exists a sequence $\{x_n\}_{n\in\mathbb{N}} \subset U$ such that

$$J_\alpha(x_n) \to m := \inf\{J_\alpha(x) \mid x \in U\}.$$

Such a sequence is called a *minimizing sequence*. Note that the convergence $\{J_\alpha(x_n)\}_{n\in\mathbb{N}}$ does not imply the convergence of $\{x_n\}_{n\in\mathbb{N}}$.

However, since convergent sequences are bounded, there exists an $M > 0$ such that

$$\frac{1}{2}\|F(x_n) - y\|_Y^2 + \frac{\alpha}{2}\|x_n - x_0\|_X^2 = J_\alpha(x_n) \leq M \qquad \text{for all } n \in \mathbb{N}. \tag{10.2}$$

It follows that

$$\frac{\alpha}{2}\left(\|x_n\|_X - \|x_0\|_X\right)^2 \le \frac{\alpha}{2}\|x_n - x_0\|_X^2 \le J_\alpha(x_n) \le M,$$

i.e., the sequence $\{x_n\}_{n\in\mathbb{N}}$ is bounded and thus contains a weakly convergent subsequence – which we again denote by $\{x_k\}_{k\in\mathbb{N}}$ for simplicity – with limit $\bar{x} \in U$ (since U is assumed to be weakly closed). This limit is a candidate for a minimizer.

Similarly, (10.2) implies that $\{F(x_k)\}_{k\in\mathbb{N}}$ is bounded in Y. By passing to a further subsequence (which we still denote by $\{x_k\}_{k\in\mathbb{N}}$), we thus obtain that $F(x_k) \rightharpoonup \bar{y} \in Y$, and the weak closedness of F yields that $\bar{y} = F(\bar{x})$. Together with the weak lower semicontinuity of norms, we obtain that

$$\begin{aligned}\frac{1}{2}\|F(\bar{x}) - y\|_Y^2 + \frac{\alpha}{2}\|\bar{x} - x_0\|_X^2 &\le \liminf_{k\to\infty} \frac{1}{2}\|F(x_k) - y\|_Y^2 + \liminf_{k\to\infty} \frac{\alpha}{2}\|x_k - x_0\|_X^2 \\ &\le \limsup_{k\to\infty}\left(\frac{1}{2}\|F(x_k) - y\|_Y^2 + \frac{\alpha}{2}\|x_k - x_0\|_X^2\right).\end{aligned}$$

By definition of the minimizing sequence, $J_\alpha(x_k) \to m$ for any subsequence as well, and hence

$$\inf_{x\in U} J_\alpha(x) \le J_\alpha(\bar{x}) \le \limsup_{k\to\infty} J_\alpha(x_k) = m = \inf_{x\in U} J_\alpha(x).$$

The infimum is thus attained in $\bar{x}$, i.e., $J_\alpha(\bar{x}) = \min_{x\in U} J_\alpha(x)$. □

Due to the nonlinearity of F, we can in general not expect the minimizer to be unique, so that we cannot introduce a well-defined mapping $y \mapsto x_\alpha$ as a regularization operator. In place of the continuity of R_α, we can therefore only show the following weaker stability result.

Theorem 10.2. *Let $F : U \to Y$ be weakly closed, $\alpha > 0$, $x_0 \in X$, and $y \in Y$. Let $\{y_n\}_{n\in\mathbb{N}}$ be a sequence with $y_n \to y$ and let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence of minimizers of J_α for y_n in place of y. Then the sequence $\{x_n\}_{n\in\mathbb{N}}$ contains a weakly convergent subsequence, and every weak cluster point of $\{x_n\}_{n\in\mathbb{N}}$ is a minimizer of J_α.*

If J_α has for y a unique minimizer, then the whole sequence converges strongly.

Proof. First, Theorem 10.1 ensures that for every $y_n \in Y$ there exists a minimizer $x_n \in U$. The minimizing property of x_n then implies for all $n \in \mathbb{N}$ and any $x \in U$ that

$$\frac{1}{2}\|F(x_n) - y_n\|_Y^2 + \frac{\alpha}{2}\|x_n - x_0\|_X^2 \le \frac{1}{2}\|F(x) - y_n\|_Y^2 + \frac{\alpha}{2}\|x - x_0\|_X^2.$$

Since $y_n \to y$, the right-hand side is bounded in $n \in \mathbb{N}$, and hence both $\{x_n\}_{n\in\mathbb{N}}$ and $\{F(x_n) - y_n\}_{n\in\mathbb{N}}$ are bounded as well. We can thus find a weakly convergent subsequence $\{x_k\}_{k\in\mathbb{N}}$ and a $\bar{x} \in U$ such that (possibly after passing to a further subsequence)

$$x_k \rightharpoonup \bar{x}, \qquad F(x_k) - y_k \rightharpoonup \bar{y}.$$

The convergence of $y_k \to y$ and the weak closedness of F then imply that $F(x_k) \rightharpoonup F(\bar{x})$. From the weak lower semicontinuity of norms, we obtain from this that

$$\frac{\alpha}{2}\|\bar{x} - x_0\|_X^2 \le \liminf_{k\to\infty} \frac{\alpha}{2}\|x_k - x_0\|_X^2, \tag{10.3}$$

$$\frac{1}{2}\|F(\bar{x}) - y\|_Y^2 \le \liminf_{k\to\infty} \frac{1}{2}\|F(x_k) - y_k\|_Y^2. \tag{10.4}$$

Using again the minimization property of the x_n, this implies that for any $x \in U$,

$$\begin{aligned} J_\alpha(\bar{x}) &= \frac{1}{2}\|F(\bar{x}) - y\|_Y^2 + \frac{\alpha}{2}\|\bar{x} - x_0\|_X^2 \\ &\le \liminf_{k\to\infty} \left(\frac{1}{2}\|F(x_k) - y_k\|_Y^2 + \frac{\alpha}{2}\|x_k - x_0\|_X^2\right) \\ &\le \limsup_{k\to\infty} \left(\frac{1}{2}\|F(x_k) - y_k\|_Y^2 + \frac{\alpha}{2}\|x_k - x_0\|_X^2\right) \\ &\le \limsup_{k\to\infty} \left(\frac{1}{2}\|F(x) - y_k\|_Y^2 + \frac{\alpha}{2}\|x - x_0\|_X^2\right) \\ &= \lim_{k\to\infty} \frac{1}{2}\|F(x) - y_k\|_Y^2 + \frac{\alpha}{2}\|x - x_0\|_X^2 \\ &= \frac{1}{2}\|F(x) - y\|_Y^2 + \frac{\alpha}{2}\|x - x_0\|_X^2 = J_\alpha(x). \end{aligned} \tag{10.5}$$

Hence $\bar{x}$ is a minimizer of J_α. Since this argument can be applied to any weakly convergent subsequence of $\{x_n\}_{n\in\mathbb{N}}$, we also obtain the second claim.

If now the minimizer x_α of J_α is unique, then every weakly convergent subsequence has the same limit, and hence the whole sequence must converge weakly to x_α. To show that this convergence is in fact strong, it suffices by (2.1) to show that $\limsup_{n\to\infty} \|x_n\|_X \le \|x_\alpha\|_X$. Assume to the contrary that this inequality does not hold. Then there must exist a subsequence $\{x_k\}_{k\in\mathbb{N}}$ with $x_k \rightharpoonup x_\alpha$ and $F(x_k) \rightharpoonup F(x_\alpha)$ but

$$\lim_{k\to\infty} \|x_k - x_0\|_X =: M > \|x_\alpha - x_0\|_X.$$

But (10.5) for $x = \bar{x} = x_\alpha$ implies that

$$\lim_{k\to\infty} \left(\frac{1}{2}\|F(x_k) - y_k\|_Y^2 + \frac{\alpha}{2}\|x_k - x_0\|_X^2\right) = \frac{1}{2}\|F(x_\alpha) - y\|_Y^2 + \frac{\alpha}{2}\|x_\alpha - x_0\|_X^2.$$

Together with the calculus for convergent sequences, this shows that

$$\begin{aligned} \lim_{k\to\infty} \frac{1}{2}\|F(x_k) - y_k\|_Y^2 &= \lim_{k\to\infty} \left(\frac{1}{2}\|F(x_k) - y_k\|_Y^2 + \frac{\alpha}{2}\|x_k - x_0\|_X^2\right) - \lim_{k\to\infty} \frac{\alpha}{2}\|x_k - x_0\|_X^2 \\ &= \frac{1}{2}\|F(x_\alpha) - y\|_Y^2 + \frac{\alpha}{2}\|x_\alpha - x_0\|_X^2 - \frac{\alpha}{2}M^2 \\ &< \frac{1}{2}\|F(x_\alpha) - y\|_Y^2 \end{aligned}$$

in contradiction to (10.4) and $\bar{x} = x_\alpha$. ◻

It remains to show that x_α converges to an x_0-minimum norm solution as $\alpha \to 0$. In contrast to the linear case, we do this already in combination with an a priori choice rule, i.e., we prove that this combination leads to a convergent regularization method. In analogy to the Section 5.2, we denote by x_α^δ a minimizer of J_α for fixed $\alpha > 0$ and noisy data $y^\delta \in Y$.

Theorem 10.3. *Let $F : U \to Y$ be weakly closed, $y \in \mathcal{R}(F)$, and $y^\delta \in B_\delta(y)$. If $\alpha(\delta)$ is a parameter choice rule such that*

$$\alpha(\delta) \to 0 \quad \text{and} \quad \frac{\delta^2}{\alpha(\delta)} \to 0 \qquad \text{for } \delta \to 0,$$

then every sequence $\{x_{\alpha(\delta_n)}^{\delta_n}\}_{n\in\mathbb{N}}$ with $\delta_n \to 0$ contains a strongly convergent subsequence, and every cluster point is an x_0-minimum norm solution of $F(x) = y$. If the x_0-minimum norm solution $x^\dagger \in U$ is unique, then the whole sequence converges strongly to $x^\dagger$.

Proof. Set $\alpha_n := \alpha(\delta_n)$ and $x_n := x_{\alpha_n}^{\delta_n}$, and let $x^\dagger$ be an x_0-minimum norm solution of $F(x) = y$. Then the minimization property of x_n implies that for all $n \in \mathbb{N}$,

$$\begin{aligned}(10.6)\qquad \frac{1}{2}\|F(x_n) - y^{\delta_n}\|_Y^2 + \frac{\alpha_n}{2}\|x_n - x_0\|_X^2 &\le \frac{1}{2}\|F(x^\dagger) - y^{\delta_n}\|_Y^2 + \frac{\alpha_n}{2}\|x^\dagger - x_0\|_X^2 \\ &\le \frac{\delta_n^2}{2} + \frac{\alpha_n}{2}\|x^\dagger - x_0\|_X^2.\end{aligned}$$

In particular,

$$(10.7)\qquad \|x_n - x_0\|_X^2 \le \frac{\delta_n^2}{\alpha_n} + \|x^\dagger - x_0\|_X^2 \qquad \text{for all } k \in \mathbb{N},$$

and the right-hand side is bounded due to the convergence $\frac{\delta_n^2}{\alpha_n} \to 0$. Hence there exists a weakly convergent subsequence $\{x_k\}_{k\in\mathbb{N}}$ and an $\bar{x} \in U$ with $x_k \rightharpoonup \bar{x}$. Similarly, we obtain from (10.6) that

$$(10.8)\qquad \frac{1}{2}\|F(x_k) - y^{\delta_k}\|_Y^2 \le \frac{\delta_k^2}{2} + \frac{\alpha_k}{2}\|x^\dagger - x_0\|_X^2 \qquad \text{for all } n \in \mathbb{N}.$$

This implies that $\{F(x_k) - y^{\delta_k}\}_{k\in\mathbb{N}}$ in turn contains a weakly convergent subsequence (which we do not further distinguish) with limit $\bar{y} \in Y$. The weak closedness of F and the strong convergence $y^{\delta_n} \to y$ then again yield that $\bar{y} = F(\bar{x}) - y$, i.e., that $F(x_k) \rightharpoonup F(\bar{x})$.

We now obtain from the weak lower semicontinuity of the norm together with (10.7) that

$$\begin{aligned}(10.9)\qquad \|\bar{x} - x_0\|_X^2 &\le \liminf_{k\to\infty} \|x_k - x_0\|_X^2 \le \limsup_{k\to\infty} \|x_k - x_0\|_X^2 \\ &\le \lim_{k\to\infty} \frac{\delta_k^2}{\alpha_k} + \|x^\dagger - x_0\|_X^2 = \|x^\dagger - x_0\|_X^2,\end{aligned}$$

and similarly from (10.8) that

$$\|F(\bar{x}) - y\|_Y^2 \le \liminf_{k\to\infty} \|F(x_k) - y^{\delta_k}\|_Y^2 \le \lim_{k\to\infty} \left(\delta_k^2 + \alpha_k\|x^\dagger - x_0\|_X^2\right) = 0.$$

Hence, $F(\bar{x}) = y$ and

$$\|\bar{x} - x_0\|_X \le \|x^\dagger - x_0\|_X = \min\{\|x - x_0\|_X \mid F(x) = y\} \le \|\bar{x} - x_0\|_X,$$

i.e., $\bar{x}$ is an x_0-minimum norm solution.

It remains to show that the subsequence $\{x_k\}_{k\in\mathbb{N}}$ converges strongly. We start from the Pythagoras identity

$$\|x_k - \bar{x}\|_X^2 = \|x_k - x_0\|_X^2 - 2\,(x_k - x_0 \mid \bar{x} - x_0)_X + \|\bar{x} - x_0\|_X^2.$$

The weak convergence $x_k \rightharpoonup \bar{x}$ then implies that

$$\lim_{k\to\infty} 2\,(x_k - x_0 \mid \bar{x} - x_0)_X = 2\,(\bar{x} - x_0 \mid \bar{x} - x_0)_X = 2\|\bar{x} - x_0\|^2.$$

Furthermore, (10.9) and the fact that both $\bar{x}$ and $x^\dagger$ are x_0-minimum norm solutions imply that

$$\lim_{k\to\infty} \|x_k - x_0\|_X = \|\bar{x} - x_0\|_X = \|x^\dagger - x_0\|_X.$$

Together, we obtain that

$$0 \le \limsup_{k\to\infty} \|x_k - \bar{x}\|_X^2 \le \|\bar{x} - x_0\|_X^2 - 2\|\bar{x} - x_0\|^2 + \|\bar{x} - x_0\|_X^2 = 0,$$

i.e., that $x_k \to \bar{x}$. The claim for a unique x_0-minimum norm solution again follows from a subsequence-subsequence argument. □

We now derive error estimates under a source conditions, where we restrict ourselves to the simplest case that corresponds to the choice $\nu = 1$ for linear inverse problems. As a motivation, we again consider the formal limit problem (6.8) for $\alpha = 0$, which in the nonlinear case becomes

$$\min_{x\in U,\, F(x)=y} \frac{1}{2}\|x - x_0\|_X^2$$

and again characterizes the x_0-minimum norm solutions. As before, we introduce a Lagrange multiplier $p \in Y$ for the equality constraint to obtain the saddle-point problem

$$\min_{x\in U}\max_{p\in Y} L(x,p), \qquad L(x,p) := \frac{1}{2}\|x - x_0\|_X^2 - (p \mid F(x) - y)_Y\,.$$

Setting the partial Fréchet derivative $L_p'(\bar{x},\bar{p})$ of L with respect to p to zero again yields the necessary condition $F(\bar{x}) = y$ for a saddle point $(\bar{x},\bar{p}) \in U \times Y$. If we assume for simplicity that the x_0-minimum norm solution $x^\dagger$ is an interior point of U, then we can also set the Fréchet derivative $L'_x(x^\dagger, p^\dagger)$ of L with respect to x in the corresponding saddle point $(x^\dagger, p^\dagger)$ to zero; this implies for all $h \in X$ that

$$0 = L'_x(x^\dagger, p^\dagger)h = \left(x^\dagger - x_0 \,\middle|\, h\right)_X - \left(p^\dagger \,\middle|\, F'(x^\dagger)h\right)_Y = \left(x^\dagger - x_0 - F'(x^\dagger)^* p^\dagger \,\middle|\, h\right)_Y,$$

i.e., the existence of a $p^\dagger \in Y$ with

$$x^\dagger - x_0 = F'(x^\dagger)^* p^\dagger.$$

This is our source condition in the nonlinear setting. However, as in the last chapter we require an additional nonlinearity condition for F in the x_0-minimum norm solution; here we assume the Lipschitz condition (9.4).

Theorem 10.4. *Let $F : U \to Y$ be Fréchet differentiable with convex domain $\operatorname{dom} F = U$. Let further $y \in \mathcal{R}(F)$ and $y^\delta \in B_\delta(y)$, and let $x^\dagger$ be an x_0-minimum norm solution such that*

(i) F' is Lipschitz continuous near $x^\dagger$ with Lipschitz constant L;

(ii) there exists a $w \in Y$ with $x^\dagger - x_0 = F'(x^\dagger)^* w$ and $L\|w\|_Y < 1$.

If $\alpha(\delta)$ is a parameter choice rule with

$$c\delta \le \alpha(\delta) \le C\delta \qquad \text{for } c, C > 0,$$

then there exist constants $c_1, c_2 > 0$ such that for all $\delta > 0$ small enough,

$$\|x^\delta_{\alpha(\delta)} - x^\dagger\|_X \le c_1\sqrt{\delta}, \tag{10.10}$$

$$\|F(x^\delta_{\alpha(\delta)}) - y^\delta\|_Y \le c_2\delta. \tag{10.11}$$

Proof. First, the minimizing property of x^δ_α for $\alpha := \alpha(\delta)$ again implies that

$$\frac{1}{2}\|F(x^\delta_\alpha) - y^\delta\|_Y^2 + \frac{\alpha}{2}\|x^\delta_\alpha - x_0\|_X^2 \le \frac{\delta^2}{2} + \frac{\alpha}{2}\|x^\dagger - x_0\|_X^2. \tag{10.12}$$

To obtain from this an estimate of $x^\delta_\alpha - x^\dagger$, we use the productive zero $x^\dagger - x^\dagger$ on the left-hand side and the Pythagoras identity, which yields the inequality

$$\|x^\delta_\alpha - x_0\|_X^2 = \|x^\delta_\alpha - x^\dagger\|_X^2 + 2\left(x^\delta_\alpha - x^\dagger \,\middle|\, x^\dagger - x_0\right)_X + \|x^\dagger - x_0\|_X^2.$$

Inserting this into (10.12) and using the source condition (ii) then shows that

$$\begin{aligned} \frac{1}{2}\|F(x^\delta_\alpha) - y^\delta\|_Y^2 + \frac{\alpha}{2}\|x^\delta_\alpha - x^\dagger\|_X^2 &\le \frac{\delta^2}{2} + \alpha\left(x^\dagger - x_0 \,\middle|\, x^\dagger - x^\delta_\alpha\right)_X \\ &= \frac{\delta^2}{2} + \alpha\left(w \,\middle|\, F'(x^\dagger)(x^\dagger - x^\delta_\alpha)\right)_Y \\ &\le \frac{\delta^2}{2} + \alpha\|w\|_Y\|F'(x^\dagger)(x^\dagger - x^\delta_\alpha)\|_Y. \end{aligned} \tag{10.13}$$

Since $x_\alpha^\delta, x^\dagger \in U$ and is U convex, the condition (i) allows us to apply Lemma 9.5 for $x = x_\alpha^\delta$ and $h = x^\dagger - x_\alpha^\delta \in U$ to obtain

$$\|F(x^\dagger) - F(x_\alpha^\delta) - F'(x^\dagger)(x^\dagger - x_\alpha^\delta)\|_Y \le \frac{L}{2}\|x^\dagger - x_\alpha^\delta\|_X^2.$$

Together with the triangle inequalities, we arrive at

$$\begin{aligned} \|F'(x^\dagger)(x^\dagger - x_\alpha^\delta)\|_Y &\le \frac{L}{2}\|x^\dagger - x_\alpha^\delta\|_X^2 + \|F(x_\alpha^\delta) - F(x^\dagger)\|_Y \\ &\le \frac{L}{2}\|x^\dagger - x_\alpha^\delta\|_X^2 + \|F(x_\alpha^\delta) - y^\delta\|_Y + \delta. \end{aligned} \tag{10.14}$$

Inserting this into (10.13) then yields that

$$\|F(x_\alpha^\delta) - y^\delta\|_Y^2 + \alpha\|x_\alpha^\delta - x^\dagger\|_X^2 \le \delta^2 + \alpha\|w\|_Y \left(L\|x^\dagger - x_\alpha^\delta\|_X^2 + 2\|F(x_\alpha^\delta) - y^\delta\|_Y + 2\delta\right).$$

We now add $\alpha^2\|w\|_Y^2$ to both sides and rearrange to obtain the inequality

$$\left(\|F(x_\alpha^\delta) - y^\delta\|_Y - \alpha\|w\|_Y\right)^2 + \alpha(1 - L\|w\|_Y)\|x_\alpha^\delta - x^\dagger\|_X^2 \le (\delta + \alpha\|w\|_Y)^2.$$

Dropping one of the two terms on the left-hand side and applying the parameter choice rule $c\delta \le \alpha \le C\delta$ then yields

$$\|F(x_\alpha^\delta) - y^\delta\|_Y \le \delta + 2\alpha\|w\|_Y \le (1 + 2C\|w\|_Y)\delta$$

as well as (since $L\|w\|_Y < 1$ by assumption)

$$\|x_\alpha^\delta - x^\dagger\|_X \le \frac{\delta + \alpha\|w\|_Y}{\sqrt{\alpha(1 - L\|w\|_Y)}} \le \frac{1 + C\|w\|_Y}{\sqrt{c(1 - L\|w\|_Y)}}\sqrt{\delta},$$

respectively, and hence the claim. □

Note that condition (ii) entails a smallness condition on $x^\dagger - x_0$: To obtain the claimed convergence rate, x_0 already has to be a sufficiently good approximation of the desired solution $x^\dagger$. Conversely, the condition indicates *which* x_0-minimum norm solution the minimizers converge to if $x^\dagger$ is not unique.

With a bit more effort, one can show analogously to Corollary 6.1 the higher rate $\delta^{\nu/(\nu+1)}$ under the stronger source condition $x^\dagger - x^0 \in \mathcal{R}((F'(x^\dagger)^*F'(x^\dagger))^{\nu/2})$ and the corresponding choice of $\alpha(\delta)$, up to the qualification $\nu_0 = 2$; see [Engl, Hanke & Neubauer 1996, Theorem 10.7].

We next consider the a posteriori choice of α according to the discrepancy principle: Set $\tau > 1$ and choose $\alpha = \alpha(\delta, y^\delta)$ such that

$$\delta < \|F(x_\alpha^\delta) - y^\delta\|_Y \le \tau\delta. \tag{10.15}$$

Theorem 10.5. *Let $F : U \to Y$ be Fréchet differentiable with convex domain* $\operatorname{dom} F = U$. *Let further $y \in \mathcal{R}(F)$ and $y^\delta \in B_\delta(y)$, and let $x^\dagger$ be an x_0-minimum norm solution such that conditions (i) and (ii) from Theorem 10.4 are satisfied. If $\alpha := \alpha(\delta, y^\delta)$ is chosen according to (10.15), then there exists a constant $c > 0$ such that*

$$\|x_\alpha^\delta - x^\dagger\|_X \le c\sqrt{\delta}.$$

Proof. From (10.15) and the minimizing property of x_α^δ, we directly obtain that

$$\frac{\delta^2}{2} + \frac{\alpha}{2}\|x_\alpha^\delta - x_0\|_X^2 < \frac{1}{2}\|F(x_\alpha^\delta) - y^\delta\|_Y^2 + \frac{\alpha}{2}\|x_\alpha^\delta - x_0\|_X^2 \le \frac{\delta^2}{2} + \frac{\alpha}{2}\|x^\dagger - x_0\|_X^2$$

and hence that

$$\frac{\alpha}{2}\|x_\alpha^\delta - x_0\|_X^2 \le \frac{\alpha}{2}\|x^\dagger - x_0\|_X^2.$$

As for (10.13) and (10.14), we can then use the conditions (i) and (ii) together with the parameter choice (10.15) to show that

$$\begin{aligned}\|x_\alpha^\delta - x^\dagger\|_X^2 &\le \|w\|_Y \left(L\|x_\alpha^\delta - x^\dagger\|_X^2 + 2\|F(x_\alpha^\delta) - y^\delta\|_Y + 2\delta\right)\\ &\le \|w\|_Y \left(L\|x_\alpha^\delta - x^\dagger\|_X^2 + 2(1+\tau)\delta\right).\end{aligned}$$

Since $L\|w\|_X < 1$, we can again rearrange this to

$$\|x_\alpha^\delta - x^\dagger\|_X^2 \le \frac{2(1+\tau)\|w\|_Y}{1 - L\|w\|_Y}\delta,$$

which yields the desired estimate. □

In contrast to Tikhonov regularization of linear problems, it is however not guaranteed that an α satisfying (10.15) exists; this requires (strong) assumptions on the nonlinearity of F. Another sufficient – and more general – assumption is the uniqueness of minimizers of J_α together with a condition on x_0.

Theorem 10.6. *Assume that for fixed $y^\delta \in B_\delta(y)$ and arbitrary $\alpha > 0$, the minimizer x_α^δ of J_α is unique. If $x_0 \in U$ and $\tau > 1$ satisfy $\|F(x_0) - y^\delta\|_Y > \tau\delta$, then there exists an $\alpha > 0$ such that (10.15) holds.*

Proof. We first show the continuity of the value function $f(\alpha) := \|F(x_\alpha^\delta) - y^\delta\|_Y$. Let $\alpha > 0$ be arbitrary and $\{\alpha_n\}_{n\in\mathbb{N}}$ be a sequence with $\alpha_n \to \alpha$ as $n \to \infty$. Then there exist $\varepsilon > 0$ and $N \in \mathbb{N}$ such that $0 < \alpha - \varepsilon \le \alpha_n \le \alpha + \varepsilon$ for all $n > N$. Let further x_α^δ be the unique minimizer of J_α and $x_n := x^\delta_{\alpha_n}$ for $n \in \mathbb{N}$ be the minimizer of J_{α_n}. The minimizing property of x_n for J_{α_n} for all $n > N$ then yields that

$$\begin{aligned}\frac{1}{2}\|F(x_n) - y^\delta\|_Y^2 + \frac{\alpha-\varepsilon}{2}\|x_n - x_0\|_X^2 &\le \frac{1}{2}\|F(x_n) - y^\delta\|_Y^2 + \frac{\alpha_n}{2}\|x_n - x_0\|_X^2 \\ &\le \frac{1}{2}\|F(x_\alpha^\delta) - y^\delta\|_Y^2 + \frac{\alpha_n}{2}\|x_\alpha^\delta - x_0\|_X^2 \\ &\le \frac{1}{2}\|F(x_\alpha^\delta) - y^\delta\|_Y^2 + \frac{\alpha+\varepsilon}{2}\|x_\alpha^\delta - x_0\|_X^2,\end{aligned}$$

which implies that both $\{x_n\}_{n>N}$ and $\{F(x_n)\}_{n>N}$ are bounded. As in the proof of Theorem 10.2, we obtain from this that

$$\lim_{n\to\infty}\left(\frac{1}{2}\|F(x_n) - y^\delta\|_Y^2 + \frac{\alpha_n}{2}\|x_n - x_0\|_X^2\right) = \frac{1}{2}\|F(x_\alpha^\delta) - y^\delta\|_Y^2 + \frac{\alpha}{2}\|x_\alpha^\delta - x_0\|_X^2 \tag{10.16}$$

as well as that (using the uniqueness of the minimizers) $x_n \to x_\alpha^\delta$. Hence $\alpha \mapsto x_\alpha^\delta$ is continuous. Together with the continuity of the norm, this implies the continuity of $g : \alpha \mapsto \frac{\alpha}{2}\|x_\alpha^\delta - x_0\|_X^2$ and thus by (10.16) also of f.

As in Lemma 6.6, we can now use the minimizing property of x_α^δ to show the monotonicity of f, which implies that

$$\begin{aligned}&\lim_{\alpha\to\infty}\|F(x_\alpha^\delta) - y^\delta\|_Y = \|F(x_0) - y^\delta\|_Y > \tau\delta, \\ &\lim_{\alpha\to 0}\|F(x_\alpha^\delta) - y^\delta\|_Y = \inf_{x\in U}\|F(x) - y^\delta\|_Y \le \|F(x^\dagger) - y^\delta\|_Y \le \delta.\end{aligned}$$

Hence, the continuous function $f(\alpha)$ attains all values in $(\delta, \tau\delta]$; in particular, there exists an α such that (10.15) holds. □

Since under our assumptions J_α is a differentiable nonlinear functional, the minimizer x_α^δ can be computed by standard methods from nonlinear optimization such as gradient or (quasi-)Newton methods. Here again the possible non-uniqueness of minimizers leads to practical difficulties. Note in particular that all claims have been about *global* minimizers of the Tikhonov functional, while (gradient-based) numerical methods in general can only produce (approximations of) *local* minimizers. This gap between theory and practice is still an open problem in inverse problems.

In the proof of Theorem 10.4, we have used the source and nonlinearity conditions to bound the right-hand side of (10.13) by suitable function of the terms on the left-hand side. It is possible to take this estimate directly as a source condition without introducing source representations or Lipschitz constants. In recent years, such *variational source conditions* have attracted increasing interest. In our context, they have the following form: There exist $\beta_1 \in [0, 1)$ and $\beta_2 \ge 0$ such that

$$\left(x^\dagger - x_0 \,\middle|\, x^\dagger - x\right)_X \le \beta_1\left(\frac{1}{2}\|x - x^\dagger\|_X^2\right) + \beta_2\|F(x) - F(x^\dagger)\|_Y \quad \text{for all } x \in U, \tag{10.17}$$

where U is a sufficiently large neighborhood of $x^\dagger$ (in particular, containing all minimizers x_α^δ of J_α). Note the different powers on the left- and right-hand sides, which are supposed to account for the different convergence speeds of error and residual.

Theorem 10.7. *Let $y \in \mathcal{R}(F)$, $y^\delta \in B_\delta(y)$, and $x^\dagger$ be a x_0-minimum norm solution satisfying the variational source condition (10.17) for some $\beta_1 < 1$. If $\alpha(\delta)$ is a parameter choice rule with*

$$c\delta \le \alpha(\delta) \le C\delta \qquad \text{for } c, C > 0,$$

then there exist constants $c_1, c_2 > 0$ such that

$$\|x^\delta_{\alpha(\delta)} - x^\dagger\|_X \le c_1\sqrt{\delta}, \tag{10.18}$$

$$\|F(x^\delta_{\alpha(\delta)}) - y^\delta\|_X \le c_2\delta. \tag{10.19}$$

Proof. From the minimizing property of x_α^δ, we again obtain the first inequality of (10.13). We now estimate this further using the variational source condition, the triangle inequality, the generalized Young inequality $ab \le \frac{1}{2\varepsilon}a^2 + \frac{\varepsilon}{2}b^2$ for $\varepsilon = \frac{1}{2}$, and the parameter choice to obtain that

$$\begin{aligned}
\frac{1}{2}\|F(x_\alpha^\delta) - y^\delta\|_Y^2 + \frac{\alpha}{2}\|x_\alpha^\delta - x^\dagger\|_X^2 &\le \frac{\delta^2}{2} + \alpha\left(x^\dagger - x_0 \,\middle|\, x^\dagger - x_\alpha^\delta\right)_X \\
&\le \frac{\delta^2}{2} + \alpha\beta_1\left(\tfrac{1}{2}\|x_\alpha^\delta - x^\dagger\|_X^2\right) + \alpha\beta_2\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y \\
&\le \frac{\delta^2}{2} + \frac{\alpha}{2}\beta_1\|x_\alpha^\delta - x^\dagger\|_X^2 + \alpha\beta_2\left(\|F(x_\alpha^\delta) - y^\delta\|_Y + \delta\right) \\
&\le \frac{\delta^2}{2} + \frac{\alpha}{2}\beta_1\|x_\alpha^\delta - x^\dagger\|_X^2 + \alpha^2\beta_2^2 + \frac{1}{4}\|F(x_\alpha^\delta) - y^\delta\|_Y^2 \\
&\quad + \alpha\beta_2\delta \\
&\le \left(\frac{1}{2} + C^2\beta_2^2 + C\beta_2\right)\delta^2 + \frac{\alpha}{2}\beta_1\|x_\alpha^\delta - x^\dagger\|_X^2 \\
&\quad + \frac{1}{4}\|F(x_\alpha^\delta) - y^\delta\|_Y^2.
\end{aligned}$$

Due to the assumption that $\beta_1 < 1$, we can absorb the last two terms on the right-hand side into the left-hand side, which yields

$$\|x_\alpha^\delta - x^\dagger\|_X \le \sqrt{\frac{1 + 2C\beta_2 + 2C^2\beta_2^2}{c(1-\beta_1)}}\sqrt{\delta} \tag{10.20}$$

as well as

$$\|F(x_\alpha^\delta) - y^\delta\|_Y \le \sqrt{2 + 4C\beta_2 + 4C^2\beta_2^2}\,\delta. \qquad \square$$

We finally study the connection between variational and classical source conditions.

Lemma 10.8. *Let $F : U \to Y$ be Fréchet differentiable and $x^\dagger$ be an x_0-minimum norm solution. If there exists a $w \in Y$ with $x^\dagger - x_0 = F'(x^\dagger)^* w$ and either*

(i) F' is Lipschitz continuous with constant $L\|w\|_Y < 1$ or

(ii) the tangential cone condition (9.6) is satisfied,

then the variational source condition (10.17) holds.

Proof. We first apply the classical source condition to the left-hand side of (10.17) and estimate

$$\begin{aligned}\left(x^\dagger - x_0 \,\middle|\, x^\dagger - x\right)_X &= \left(F'(x^\dagger)^* w \,\middle|\, x^\dagger - x\right)_X \\ &= \left(w \,\middle|\, F'(x^\dagger)(x^\dagger - x)\right)_Y \\ &\le \|w\|_Y \|F'(x^\dagger)(x^\dagger - x)\|_Y \\ &\le \|w\|_Y \left(\|F(x) - F(x^\dagger) - F'(x^\dagger)(x^\dagger - x)\|_Y + \|F(x) - F(x^\dagger)\|_Y\right).\end{aligned}$$

If now assumption (i) holds, we can apply Lemma 9.5 to obtain the inequality

$$\left(x^\dagger - x_0 \,\middle|\, x^\dagger - x\right)_X \le \|w\|_Y \left(\frac{L}{2}\|x^\dagger - x\|_X^2 + \|F(x) - F(x^\dagger)\|_Y\right),$$

i.e., (10.17) with $\beta_1 = L\|w\|_Y < 1$ and $\beta_2 = \|w\|_Y$.

On the other hand, if assumption (ii) holds, we can directly estimate

$$\left(x^\dagger - x_0 \,\middle|\, x^\dagger - x\right)_X \le \|w\|_Y (\eta + 1)\|F(x) - F(x^\dagger)\|_Y,$$

which implies (10.17) with $\beta_1 = 0$ and $\beta_2 = (1 + \eta)\|w\|_Y > 0$. □

For a linear operator $T \in \mathbb{L}(X, Y)$, we of course do not need any nonlinearity condition; in this case the variational source condition (10.17) is equivalent to the classical source condition $x^\dagger \in \mathcal{R}(T^*)$, see [Andreev et al. 2015, Lemma 2]. For nonlinear operators, however, it is a weaker (albeit even more abstract) condition. The main advantage of this type of condition is that it does not involve the Fréchet derivative of F and hence can also be applied for non-differentiable F; furthermore, it can be applied to generalized Tikhonov regularization, in particular in Banach spaces; see, e.g., [Hofmann et al. 2007; Scherzer et al. 2009; Schuster et al. 2012].

11 ITERATIVE REGULARIZATION

There also exist iterative methods for nonlinear inverse problems that, like the Landweber iteration, construct a sequence of approximations and can be combined with a suitable termination criterion to obtain a regularization method. Specifically, a *(convergent) iterative regularization method* refers to a procedure that constructs for given $y^\delta \in Y$ and $x_0 \in U$ a sequence $\{x_n^\delta\}_{n\in\mathbb{N}} \subset U$ together with a stopping index $N(\delta, y^\delta)$, such that for all $y \in \mathcal{R}(F)$ and all $x_0 = x_0^\delta$ sufficiently close to an isolated solution $x^\dagger \in U$ of $F(x) = y$, we have that[1]

$$N(0,y) < \infty, \quad x_{N(0,y)} = x^\dagger \qquad \text{or} \qquad N(0,y) = \infty, \quad x_n \to x^\dagger \text{ for } n \to \infty, \tag{11.1a}$$

$$\lim_{\delta\to 0} \sup_{y^\delta \in B_\delta(y)} \|x^\delta_{N(\delta,y^\delta)} - x^\dagger\|_X = 0. \tag{11.1b}$$

The first condition states that for exact data (i.e., $\delta = 0$), the sequence either converges to a solution or reaches one after finitely many steps. The second condition corresponds to the definition of a convergent regularization method in the linear setting.

We again terminate by the Morozov discrepancy principle: Set $\tau > 1$ and choose $N = N(\delta, y^\delta)$ such that

$$\|F(x_N^\delta) - y^\delta\|_Y \le \tau\delta < \|F(x_n^\delta) - y^\delta\|_Y \qquad \text{for all } n < N. \tag{11.2}$$

In this case, a sufficient condition for (11.1b) is the monotonicity and stability of the method. Here and in the following, we again denote by x_n the elements of the sequence generated for the exact data $y \in \mathcal{R}(F)$ and by x_n^δ the elements for the noisy data $y^\delta \in B_\delta(y)$.

Lemma 11.1. *Let $N(\delta, y^\delta)$ be chosen by the discrepancy principle (11.2). If an iterative method for a continuous operator $F : U \to Y$ satisfies the condition (11.1a) as well as*

$$\|x_n^\delta - x^\dagger\|_X \le \|x_{n-1}^\delta - x^\dagger\|_X \qquad \textit{for all } n \in \{1, \dots, N(\delta, y^\delta)\}, \tag{11.3a}$$

$$\lim_{\delta\to 0} \|x_n^\delta - x_n\|_X = 0 \qquad \textit{for every fixed } n \in \mathbb{N}, \tag{11.3b}$$

then the condition (11.1b) is also satisfied.

[1]In contrast to the previous chapters, we denote here by $x^\dagger$ not an (x_0-)minimum norm solution, but any solution of $F(x) = y$.

Proof. Let $F : U \to Y$ be continuous, $\{y^{\delta_k}\}_{k\in\mathbb{N}}$ with $y^{\delta_k} \in B_{\delta_k}(y)$ and $\delta_k \to 0$ as $k \to \infty$, and set $N_k := N(\delta_k, y^{\delta_k})$. We first consider the case that $\{N_k\}_{k\in\mathbb{N}}$ is bounded and hence that the set $\{N_k \mid k \in \mathbb{N}\} \subset \mathbb{N}$ is finite. After passing to a subsequence if necessary, we can therefore assume that $N_k = \bar N$ for all $k \in \mathbb{N}$. It then follows from (11.3b) that $x_{\bar N}^{\delta_k} \to x_{\bar N}$ as $k \to \infty$. Since all N_k are chosen according to the discrepancy principle (11.2), we have that

$$\|F(x_{\bar N}^{\delta_k}) - y^{\delta_k}\|_Y \le \tau\delta_k \qquad \text{for all } k \in \mathbb{N}.$$

Passing to the limit on both sides and using the continuity of F then yields that $F(x_{\bar N}) = y$, i.e., $x_{\bar N}^{\delta_k}$ converges to a solution of $F(x) = y$ and the condition (11.1b) is thus satisfied.

Otherwise, there exists a subsequence with $N_k \to \infty$. We can assume (possibly after passing to a further subsequence) that N_k is increasing. Then (11.3a) yields that for all $l \le k$,

$$\|x_{N_k}^{\delta_k} - x^\dagger\|_X \le \|x_{N_l}^{\delta_k} - x^\dagger\|_X \le \|x_{N_l}^{\delta_k} - x_{N_l}\|_X + \|x_{N_l} - x^\dagger\|_X.$$

Let now $\varepsilon > 0$ be arbitrary. Since we have assumed that condition (11.1a) holds, there exists an $L > 0$ such that $\|x_{N_L} - x^\dagger\|_X \le \frac{\varepsilon}{2}$. Similarly, (11.3b) for $n = N_L$ shows the existence of a $K > 0$ such that $\|x_{N_L}^{\delta_k} - x_{N_L}\|_X \le \frac{\varepsilon}{2}$ for all $k \ge K$. Hence, the condition (11.1b) holds in this case as well. □

A sequence $\{x_n\}_{n\in\mathbb{N}}$ satisfying (11.3a) is called *Féjer monotone*; this property is fundamental for the convergence proof of many iterative methods.

In general, iterative methods for nonlinear inverse problems rely on a linearization of F, with different methods applying the linearization at different points in the iteration.

11.1 LANDWEBER REGULARIZATION

Analogously to the linear Landweber regularization, we start from the characterization of the wanted solution $x^\dagger$ as a minimizer of the functional $J_0(x) = \frac{1}{2}\|F(x) - y\|_Y^2$. If F is Fréchet differentiable, the chain rule yields the necessary optimality condition

$$0 = J_0'(x^\dagger)h = \left(F(x^\dagger) - y \,\middle|\, F(x^\dagger)'h\right)_Y = \left(F'(x^\dagger)^*(F(x^\dagger) - y) \,\middle|\, h\right)_X \qquad \text{for all } h \in X.$$

This is now a nonlinear equation for $x^\dagger$, which as in the linear case can be written as a fixed-point equation. This leads to the nonlinear Richardson iteration

$$x_{n+1} = x_n - \omega_n F'(x_n)^*(F(x_n) - y),$$

for which we can expect convergence if $\omega_n\|F'(x_n)^*\|_{\mathbb{L}(Y,X)}^2 < 1$. (Alternatively, (11.1) can be interpreted as a steepest descent method with step size ω_n for the minimization of J_0.) For simplicity, we assume in the following that $\|F'(x)\|_{\mathbb{L}(X,Y)} < 1$ for all x sufficiently close to

$x^\dagger$, so that we can take $\omega_n = 1$. (This is not a significant restriction since can always scale F and y appropriately without changing the solution of $F(x) = y$.) Furthermore, we assume that F is continuously Fréchet differentiable and satisfies the tangential cone condition (9.6) in a neighborhood of $x^\dagger$. Specifically, we make the following assumption:

Assumption 11.2. Let $F : U \to Y$ be continuously differentiable and $x_0 \in U$. Assume that there exists an $r > 0$ such that

(i) $B_{2r}(x_0) \subset U$;

(ii) there exists a solution $x^\dagger \in B_r(x_0)$;

(iii) for all $x, \tilde{x} \in B_{2r}(x_0)$,

$$\|F'(x)\|_{\mathbb{L}(X,Y)} \leq 1, \tag{11.4}$$

$$\|F(x) - F(\tilde{x}) - F'(x)(x - \tilde{x})\|_Y \leq \eta \|F(x) - F(\tilde{x})\|_Y \qquad \text{with } \eta < \tfrac{1}{2}. \tag{11.5}$$

Under these assumptions, the nonlinear Landweber iteration (11.1) is well-posed and Féjer monotone even for noisy data $y^\delta \in B_\delta(y)$.

Lemma 11.3. *Let Assumption 11.2 hold. If $x_n^\delta \in B_r(x^\dagger)$ for some $\delta \geq 0$ and satisfies*

$$\|F(x_n^\delta) - y^\delta\|_Y \geq 2\frac{1+\eta}{1-2\eta}\delta, \tag{11.6}$$

then

$$\|x_{n+1}^\delta - x^\dagger\|_X \leq \|x_n^\delta - x^\dagger\|_X \tag{11.7}$$

and thus $x_{n+1}^\delta \in B_r(x^\dagger) \subset B_{2r}(x_0)$.

Proof. The iteration (11.1) together with (11.4) for $x_n^\delta \in B_r(x^\dagger) \subset B_{2r}(x_0)$ lead to the estimate

$$\begin{aligned}
\|x_{n+1}^\delta - x^\dagger\|_X^2 - \|x_n^\delta - x^\dagger\|_X^2 &= 2\left(x_{n+1}^\delta - x_n^\delta \,\middle|\, x_n^\delta - x^\dagger\right)_X + \|x_{n+1}^\delta - x_n^\delta\|_X^2 \\
&= 2\left(F'(x_n^\delta)^*(y^\delta - F(x_n^\delta)) \,\middle|\, x_n^\delta - x^\dagger\right)_X \\
&\qquad + \|F'(x_n^\delta)^*(y^\delta - F(x_n^\delta))\|_X^2 \\
&\leq 2\left(y^\delta - F(x_n^\delta) \,\middle|\, F'(x_n^\delta)(x_n^\delta - x^\dagger)\right)_Y + \|y^\delta - F(x_n^\delta)\|_Y^2 \\
&= 2\left(y^\delta - F(x_n^\delta) \,\middle|\, y^\delta - F(x_n^\delta) + F'(x_n^\delta)(x_n^\delta - x^\dagger)\right)_Y \\
&\qquad - \|y^\delta - F(x_n^\delta)\|_Y^2 \\
&\leq \|y^\delta - F(x_n^\delta)\|_Y \big(2\|y^\delta - F(x_n^\delta) + F'(x_n^\delta)(x_n^\delta - x^\dagger)\|_Y \\
&\qquad - \|y^\delta - F(x_n^\delta)\|_Y\big).
\end{aligned}$$

Inserting the productive zero $F(x^\dagger) - y$ in the first norm inside the parentheses and applying the triangle inequality as well as the tangential cone condition (11.5) then yields that

$$\begin{aligned}\|y^\delta - F(x_n^\delta) + F'(x_n^\delta)(x_n^\delta - x^\dagger)\|_Y &\leq \delta + \|F(x_n^\delta) - F(x^\dagger) - F'(x_n^\delta)(x_n^\delta - x^\dagger)\|_Y \\ &\leq \delta + \eta\|F(x_n^\delta) - F(x^\dagger)\|_Y \\ &\leq (1+\eta)\delta + \eta\|F(x_n^\delta) - y^\delta\|_Y\end{aligned}$$

and hence that

$$\|x_{n+1}^\delta - x^\dagger\|_X^2 - \|x_n^\delta - x^\dagger\|_X^2 \leq \|y^\delta - F(x_n^\delta)\|_Y\big(2(1+\eta)\delta - (1-2\eta)\|y^\delta - F(x_n^\delta)\|_Y\big). \tag{11.8}$$

By (11.6), the term in parentheses is non-positive, from which the desired monotonicity follows. □

By induction, this shows that $x_n^\delta \in B_{2r}(x_0) \subset U$ as long as (11.6) holds. If we choose τ for the discrepancy principle (11.2) such that

$$\tau > 2\frac{1+\eta}{1-2\eta} > 2, \tag{11.9}$$

then this is the case for all $n \leq N(\delta, y^\delta)$. This choice also guarantees that the stopping index $N(\delta, y^\delta)$ is finite.

Theorem 11.4. *Let Assumption 11.2 hold. If $N(\delta, y^\delta)$ is chosen according to the discrepancy principle (11.2) with τ satisfying (11.9) then*

$$N(\delta, y^\delta) < C\delta^{-2} \qquad \text{for some } C > 0. \tag{11.10}$$

For exact data (i.e., $\delta = 0$),

$$\sum_{n=0}^{\infty} \|F(x_n) - y\|_Y^2 < \infty. \tag{11.11}$$

Proof. Since $x_0^\delta = x_0 \in B_{2r}(x_0)$ and by the choice of τ, we can apply Lemma 11.3 for all $n < N = N(\delta, y^\delta)$. In particular, it follows from (11.8) and (11.9) that

$$\|x_{n+1}^\delta - x^\dagger\|_X^2 - \|x_n^\delta - x^\dagger\|_X^2 < \|y^\delta - F(x_n^\delta)\|_Y^2 \left(\frac{2}{\tau}(1+\eta) + 2\eta - 1\right) \qquad \text{for all } n < N.$$

Summing from $n = 0$ to $N - 1$ and telescoping thus yields

$$\left(1 - 2\eta - \frac{2}{\tau}(1+\eta)\right)\sum_{n=0}^{N-1} \|F(x_n^\delta) - y^\delta\|_Y^2 < \|x_0 - x^\dagger\|_X^2 - \|x_N^\delta - x^\dagger\|_X^2 \leq \|x_0 - x^\dagger\|_X^2.$$

Since N is chosen according to the discrepancy principle, we have that $\|F(x_n^\delta) - y^\delta\|_Y > \tau\delta$ for all $n < N$. Together we thus obtain that

$$N\tau^2\delta^2 < \sum_{n=0}^{N-1} \|F(x_n^\delta) - y^\delta\|_Y^2 < \left(1 - 2\eta - 2\tau^{-1}(1+\eta)\right)^{-1} \|x_0 - x^\dagger\|_X^2$$

and hence (11.10) for $C := \left((1-2\eta)\tau^2 - 2(1+\eta)\tau\right)^{-1} \|x_0 - x^\dagger\|_X^2 > 0$.

For $\delta = 0$, (11.6) is satisfied for all $n \in \mathbb{N}$, and obtain directly from (11.8) by summing and telescoping that

$$(1-2\eta)\sum_{n=0}^{N-1} \|F(x_n) - y\|_Y^2 \leq \|x_0 - x^\dagger\|_X^2 \qquad \text{for all } N \in \mathbb{N}.$$

Passing to the limit $N \to \infty$ then yields (11.11). □

Although (11.11) implies that $F(x_n) \to y$ for exact data $y \in \mathcal{R}(F)$, we cannot yet conclude that the x_n converge. This we show next.

Theorem 11.5. *Let Assumption 11.2 hold. Then $x_n \to \bar{x}$ with $F(\bar{x}) = y$ as $n \to \infty$.*

Proof. We show that $\{e_n\}_{n\in\mathbb{N}}$ with $e_n := x_n - x^\dagger$ is a Cauchy sequence. Let $m, n \in \mathbb{N}$ with $m \geq n$ be given and choose $k \in \mathbb{N}$ with $m \geq k \geq n$ such that

$$\|y - F(x_k)\|_Y \leq \|y - F(x_j)\|_Y \qquad \text{for all } n \leq j \leq m. \tag{11.12}$$

(I.e., we chose $k \in \{n, \dots, m\}$ such that the residual – which need not be monotone in the nonlinear case – is minimal in this range.) We now estimate

$$\|e_m - e_n\|_X \leq \|e_m - e_k\|_X + \|e_k - e_n\|_X$$

and consider each term separately. First,

$$\begin{aligned} \|e_m - e_k\|_X^2 &= 2\,(e_k - e_m \mid e_k)_X + \|e_m\|_X^2 - \|e_k\|_X^2, \\ \|e_k - e_n\|_X^2 &= 2\,(e_k - e_n \mid e_k)_X + \|e_n\|_X^2 - \|e_k\|_X^2. \end{aligned}$$

It follows from Lemma 11.3 that $\|e_n\|_X \geq 0$ is decreasing and thus converges to some $\varepsilon \geq 0$. Hence, both differences on the right-hand side converge to zero as $n \to \infty$, and it remains to look at the inner products. Here, inserting the definition of e_n, telescoping the sum, and using the iteration (11.1) yields that

$$e_m - e_k = x_m - x_k = \sum_{j=k}^{m-1} x_{j+1} - x_j = \sum_{j=k}^{m-1} F'(x_j)^*(y - F(x_j)).$$

Inserting this into the inner product, generously adding productive zeros, and using the tangential cone condition (11.5) then leads to

$$
\begin{aligned}
(e_k - e_m \mid e_k)_X &= \sum_{j=k}^{m-1} -\left(y - F(x_j) \,\middle|\, F'(x_j)(x_k - x^\dagger)\right)_Y \\
&\leq \sum_{j=k}^{m-1} \|y - F(x_j)\|_Y \|F'(x_j)(x_k - x_j + x_j - x^\dagger)\|_Y \\
&\leq \sum_{j=k}^{m-1} \|y - F(x_j)\|_Y \big(\|y - F(x_j) - F'(x_j)(x^\dagger - x_j)\|_Y + \|y - F(x_k)\|_Y \\
&\qquad\qquad + \|F(x_j) - F(x_k) - F'(x_j)(x_j - x_k)\|_Y\big) \\
&\leq (1+\eta) \sum_{j=k}^{m-1} \|y - F(x_j)\|_Y \|y - F(x_k)\|_Y + 2\eta \sum_{j=k}^{m-1} \|y - F(x_j)\|_Y^2 \\
&\leq (1+3\eta) \sum_{j=k}^{m-1} \|y - F(x_j)\|_Y^2,
\end{aligned}
$$

where we have used the definition (11.12) of k in the last estimate. Similarly we obtain that

$$
(e_k - e_n \mid e_k)_X \leq (1+3\eta) \sum_{j=n}^{k-1} \|y - F(x_j)\|_Y^2.
$$

Due to Theorem 11.4, both remainder terms converge to zero as $n \to \infty$. Hence $\{e_n\}_{n\in\mathbb{N}}$ and therefore also $\{x_n\}_{n\in\mathbb{N}}$ are Cauchy sequences, which implies that $x_n \to \bar{x}$ with $F(\bar{x}) = y$ (due to (11.11)). □

It remains to show the convergence condition (11.1b) for noisy data.

Theorem 11.6. *Let Assumption 11.2 hold. Then $x_{N(\delta,y^\delta)} \to \bar{x}$ with $F(\bar{x}) = y$ as $\delta \to 0$.*

Proof. We apply Lemma 11.1, for which we have already shown condition (11.1a) in Theorem 11.5. Since F and F' are by assumption continuous, the right-hand side of (11.1) for fixed $n \in \mathbb{N}$ depends continuously on x_n. Hence for all $k \leq n$, the right-hand side of (11.1) for x_{k+1}^δ converges to that for x_{k+1} as $\delta \to 0$, which implies the stability condition (11.3b). Finally, the monotonicity condition (11.3a) follows from Lemma 11.3, and hence Lemma 11.1 yields (11.1b). □

Under the usual source condition $x^\dagger - x_0 \in \mathcal{R}(F'(x^\dagger)^*)$ – together with additional, technical, assumptions on the nonlinearity of F – it is possible to show the expected convergence rate of $\mathcal{O}(\sqrt{\delta})$, see [Hanke, Neubauer & Scherzer 1995, Theorem 3.2], [Kaltenbacher, Neubauer & Scherzer 2008, Theorem 2.13].

11.2 LEVENBERG–MARQUARDT METHOD

As in the linear case, one drawback of the Landweber iteration is that (11.10) shows that $N(\delta, y^\delta) = O(\delta^{-2})$ may be necessary to satisfy the discrepancy principle, which in practice can be too many. Faster iterations can be built on Newton-type methods. For the original equation $F(x) = y$, one step of Newton's method consists in solving the linearized equation

$$F'(x_n)h_n = -(F(x_n) - y) \tag{11.13}$$

and setting $x_{n+1} := x_n + h_n$. However, if F is completely continuous, the Fréchet derivative $F'(x_n)$ is compact by Theorem 9.6, and hence (11.13) is in general ill-posed as well. The idea is now to apply Tikhonov regularization to the Newton step (11.13), i.e., to compute h_n as the solution of the minimization problem

$$\min_{h\in X} \frac{1}{2}\|F'(x_n)h + F(x_n) - y\|_Y^2 + \frac{\alpha_n}{2}\|h\|_X^2 \tag{11.14}$$

for suitable $\alpha_n > 0$. Using Lemma 6.3 and $h_n = x_{n+1} - x_n$, this leads to an explicit scheme that is known as the *Levenberg–Marquardt method*:

$$x_{n+1} = x_n + \left(F'(x_n)^*F'(x_n) + \alpha_n \,\mathrm{Id}\right)^{-1} F'(x_n)^*(y - F(x_n)). \tag{11.15}$$

We now show similarly to the Landweber iteration that (11.15) leads to an iterative regularization method even for noisy data $y^\delta \in B_\delta(y)$. This requires choosing α_n appropriately; we do this such that the corresponding minimizer h_{α_n} satisfies for some $\sigma \in (0,1)$ the equation

$$\|F'(x_n^\delta)h_{\alpha_n} + F(x_n^\delta) - y^\delta\|_Y = \sigma\|F(x_n^\delta) - y^\delta\|_Y. \tag{11.16}$$

Note that this is a heuristic choice rule; we thus require additional assumptions.

Assumption 11.7. Let $F : U \to Y$ be continuously differentiable and $x_0 \in U$. Assume that there exists an $r > 0$ such that

(i) $B_{2r}(x_0) \subset U$;

(ii) there exists a solution $x^\dagger \in B_r(x_0)$;

(iii) there exists a $\gamma > 1$ such that

$$\|F'(x_n^\delta)(x^\dagger - x_n^\delta) + F(x_n^\delta) - y^\delta\|_Y \le \frac{\sigma}{\gamma}\|F(x_n^\delta) - y^\delta\|_Y \qquad \text{for all } n \in \mathbb{N}. \tag{11.17}$$

Theorem 11.8. *If Assumption 11.7 holds, then there exists an $\alpha_n > 0$ satisfying (11.16).*

Proof. Set $f_n(\alpha) := \|F'(x_n^\delta)h_\alpha + F(x_n^\delta) - y^\delta\|_Y$. Since $F'(x_n^\delta)$ is linear, the minimizer h_α of (11.14) is unique for all $\alpha > 0$. As in the proof of Theorem 10.6, this implies the continuity of f_n as well as that

$$\lim_{\alpha\to\infty} f_n(\alpha) = \|F(x_n^\delta) - y^\delta\|_Y,$$
$$\lim_{\alpha\to 0} f_n(\alpha) = \inf_{h\in X} \|F'(x_n^\delta)h + F(x_n^\delta) - y^\delta\|_Y \le \|F'(x_n^\delta)(x^\dagger - x_n^\delta) + F(x_n^\delta) - y^\delta\|_Y.$$

By assumption, we now have that

$$\lim_{\alpha\to 0} f_n(\alpha) \le \frac{\sigma}{\gamma}\|F(x_n^\delta) - y^\delta\|_Y < \sigma\|F(x_n^\delta) - y^\delta\|_Y < \|F(x_n^\delta) - y^\delta\|_Y = \lim_{\alpha\to\infty} f_n(\alpha),$$

which together with the continuity of $f_n(\alpha)$ implies the existence of a solution $\alpha_n > 0$ of $f_n(\alpha) = \sigma\|F(x_n^\delta) - y^\delta\|_Y$. ☐

For this choice of of α_n, we can again show the Féjer monotonicity property (11.3a).

Lemma 11.9. *Let Assumption 11.7 hold. If $x_n^\delta \in B_r(x^\dagger)$, then*

$$\|x_n^\delta - x^\dagger\|_X^2 - \|x_{n+1}^\delta - x^\dagger\|_X^2 \ge \|x_{n+1}^\delta - x_n^\delta\|_X^2 + \frac{2(\gamma-1)\sigma^2}{\gamma\alpha_n}\|F(x_n^\delta) - y^\delta\|_Y^2. \tag{11.18}$$

In particular,

$$\|x_{n+1}^\delta - x^\dagger\|_X \le \|x_n^\delta - x^\dagger\|_X \tag{11.19}$$

and hence $x_{n+1}^\delta \in B_r(x^\dagger) \subset B_{2r}(x_0)$.

Proof. We proceed as for Lemma 11.3 by using the iteration (11.15) to estimate the error difference, this time applying the parameter choice (11.16) in place of the discrepancy principle. For the sake of legibility, we set $T_n := F'(x_n^\delta)$, $h_n := x_{n+1}^\delta - x_n^\delta$, and $\tilde y_n := y^\delta - F(x_n^\delta)$. First, we rewrite (11.15) as $\alpha_n h_n = T_n^*\tilde y_n - T_n^*T_n h_n$, which implies that

$$\left(x_{n+1}^\delta - x_n^\delta \,\middle|\, x_n^\delta - x^\dagger\right)_X = \alpha_n^{-1}\left(\tilde y_n - T_n h_n \,\middle|\, T_n(x_n^\delta - x^\dagger)\right)_Y \tag{11.20}$$

and similarly that

$$\left(x_{n+1}^\delta - x_n^\delta \,\middle|\, x_{n+1}^\delta - x_n^\delta\right)_X = \alpha_n^{-1}\left(\tilde y_n - T_n h_n \,\middle|\, T_n h_n\right)_Y. \tag{11.21}$$

Together with the productive zero $\tilde{y}_n - \tilde{y}_n$, this shows that

$$
\begin{aligned}
\|x^\delta_{n+1} - x^\dagger\|_X^2 - \|x_n - x^\dagger\|_X^2 &= 2\left(x^\delta_{n+1} - x^\delta_n \,\middle|\, x^\delta_n - x^\dagger\right)_X + \|x^\delta_{n+1} - x^\delta_n\|_X^2 \\
&= 2\alpha_n^{-1}\left(\tilde{y}_n - T_n h_n \,\middle|\, \tilde{y}_n + T_n(x^\delta_n - x^\dagger)\right)_Y \\
&\quad + 2\alpha_n^{-1}\left(\tilde{y}_n - T_n h_n \mid T_n h_n - \tilde{y}_n\right)_Y - \|x^\delta_{n+1} - x^\delta_n\|_X^2 \\
&= 2\alpha_n^{-1}\left(\tilde{y}_n - T_n h_n \,\middle|\, \tilde{y}_n - T_n(x^\dagger - x^\delta_n)\right)_Y \\
&\quad - 2\alpha_n^{-1}\|\tilde{y}_n - T_n h_n\|_Y^2 - \|x^\delta_{n+1} - x^\delta_n\|_X^2 \\
&\leq 2\alpha_n^{-1}\|\tilde{y}_n - T_n h_n\|_Y \|\tilde{y}_n - T_n(x^\dagger - x^\delta_n)\|_Y \\
&\quad - 2\alpha_n^{-1}\|\tilde{y}_n - T_n h_n\|_Y^2 - \|x^\delta_{n+1} - x^\delta_n\|_X^2.
\end{aligned}
$$

For the terms with h_n, we can directly insert the parameter choice rule (11.16). For the terms with $x^\dagger$, we apply the assumption (11.17) together with (11.16) to obtain that

$$\|\tilde{y}_n - T_n(x^\dagger - x^\delta_n)\|_Y \leq \frac{\sigma}{\gamma}\|\tilde{y}_n\|_Y = \frac{1}{\gamma}\|\tilde{y}_n - T_n h_n\|_Y.$$

Inserting this, rearranging, and multiplying with -1 now yields (11.18). □

We next show that for noisy data $y^\delta \in B_\delta(y)$, the discrepancy principle (11.2) yields a finite stopping criterion $N(\delta, y^\delta)$. This requires a stronger version of the tangential cone condition (11.5).

Assumption 11.10. Let Assumption 11.7 hold with (iii) replaced by

(iii′) there exist $M > 0$ and $c > 0$ such that for all $x, \tilde{x} \in B_{2r}(x_0)$,

$$\|F'(x)\|_{\mathbb{L}(X,Y)} \leq M, \tag{11.22}$$

$$\|F(x) - F(\tilde{x}) - F'(x)(x - \tilde{x})\|_Y \leq c\|x - \tilde{x}\|_X \|F(x) - F(\tilde{x})\|_Y. \tag{11.23}$$

Theorem 11.11. *Let Assumption 11.10 hold. If $N(\delta, y^\delta)$ is chosen according to the discrepancy principle (11.2) with $\tau > \sigma^{-1}$ and if $\|x_0 - x^\dagger\|_X$ is sufficiently small, then*

$$N(\delta, y^\delta) < C(1 + |\log \delta|) \qquad \textit{for some } C > 0.$$

Proof. We first show that under these assumptions, the error decreases up to the stopping index. Assume that $N := N(\delta, y^\delta) \geq 1$ (otherwise there is nothing to show) and that

$$\|x_0 - x^\dagger\|_X \leq \min\{r, \tilde{r}\}, \qquad \tilde{r} := \frac{\sigma\tau - 1}{c(1 + \tau)}. \tag{11.24}$$

From (11.23) with $x = x_0$ and $\tilde{x} = x^\dagger$, we then obtain by inserting $y - y$ that

$$\begin{aligned}\|F'(x_0)(x^\dagger - x_0) + F(x_0) - y^\delta\|_Y &\leq \delta + \|F(x_0) - y - F'(x_0)(x_0 - x^\dagger)\|_Y \\ &\leq \delta + c\|x_0 - x^\dagger\|_X\|F(x_0) - y\|_Y \\ &\leq (1 + c\|x_0 - x^\dagger\|_X)\delta + c\|x_0 - x^\dagger\|_X\|F(x_0) - y^\delta\|_Y.\end{aligned}$$

Since x_0 by assumption does not satisfy the discrepancy principle, $\delta < \tau^{-1}\|F(x_0) - y^\delta\|_Y$. Inserting this thus yields (11.17) with $\gamma := \sigma\tau(1 + c(1 + \tau)\|x_0 - x^\dagger\|_X)^{-1} > 1$ for $\|x_0 - x^\dagger\|_X$ sufficiently small. Hence Lemma 11.9 implies that

$$\|x_1^\delta - x^\dagger\|_X \leq \|x_0 - x^\dagger\|_X \leq \min\{r, \tilde{r}\}$$

and therefore in particular that $x_1^\delta \in B_{2r}(x_0) \subset U$. If now $N > 1$, we obtain as above that

$$\begin{aligned}\|F'(x_1^\delta)(x^\dagger - x_1^\delta) + F(x_1^\delta) - y^\delta\|_Y &\leq (1 + c\|x_1^\delta - x^\dagger\|_X)\delta + c\|x_1^\delta - x^\dagger\|_X\|F(x_1^\delta) - y^\delta\|_Y \\ &\leq (1 + c\|x_0 - x^\dagger\|_X)\delta + c\|x_0 - x^\dagger\|_X\|F(x_1^\delta) - y^\delta\|_Y.\end{aligned}$$

By induction, the iteration (11.15) is thus well-defined for all $n < N$, and (11.18) holds.

Proceeding as for the Landweber iteration by summing the residuals now requires a uniform bound on α_n. For this, we use that with T_n, h_n and $\tilde{y}_n$ as in the proof of Lemma 11.9,

$$(T_nT_n^* + \alpha_n \operatorname{Id})\left(\tilde{y}_n - T_nh_n\right) = T_n\left(T_n^*\tilde{y}_n - T_n^*T_nh_n - \alpha_nh_n\right) + \alpha_n\tilde{y}_n = \alpha_n\tilde{y}_n,$$

where we have used the iteration (11.15) in the last step. Using the assumption $\|T_n\|_{\mathbb{L}(X,Y)} \leq M$ and the parameter choice (11.16) then implies that

$$\begin{aligned}\alpha_n\|\tilde{y}_n\|_Y &= \|(T_nT_n^* + \alpha_n \operatorname{Id})(\tilde{y}_n - T_nh_n)\|_Y \\ &\leq (M^2 + \alpha_n)\|\tilde{y}_n - T_nh_n\|_Y \\ &= (M^2 + \alpha_n)\sigma\|\tilde{y}_n\|_Y.\end{aligned} \tag{11.25}$$

Solving (11.25) for α_n now yields that $\alpha_n \leq \frac{\sigma M^2}{1-\sigma}$, which together with (11.18) leads to

$$\|x_n^\delta - x^\dagger\|_X^2 - \|x_{n+1}^\delta - x^\dagger\|_X^2 \geq \frac{2(\gamma - 1)(1 - \sigma)\sigma}{\gamma M^2}\|F(x_n^\delta) - y^\delta\|_Y^2 \qquad \text{for all } n < N.$$

Since N was chosen according to discrepancy principle (11.2), we can sum this inequality from $n = 0$ to $N - 1$ to obtain the estimate

$$N(\tau\delta)^2 \leq \sum_{n=0}^{N-1}\|F(x_n^\delta) - y^\delta\|_Y^2 \leq \frac{\gamma M^2}{2(\gamma - 1)(1 - \sigma)\sigma}\|x_0 - x^\dagger\|_X.$$

This implies that N is finite for all $\delta > 0$.

For the logarithmic estimate, we use the parameter choice (11.16) together with the assumption (11.23) to show that for arbitrary $n < N$,

$$\begin{aligned}\sigma\|F(x_n^\delta) - y^\delta\|_Y &= \|F'(x_n^\delta)h_n + F(x_n^\delta) - y^\delta\|_Y \\ &\geq \|F(x_{n+1}^\delta) - y^\delta\|_Y - \|F'(x_n^\delta)h_n + F(x_n^\delta) - F(x_{n+1}^\delta)\|_Y \\ &\geq \|F(x_{n+1}^\delta) - y^\delta\|_Y - c\|h_n\|_X\|F(x_{n+1}^\delta) - F(x_n^\delta)\|_Y \\ &\geq (1 - c\|h_n\|_X)\|F(x_{n+1}^\delta) - y^\delta\|_Y - c\|h_n\|_X\|F(x_n^\delta) - y^\delta\|_Y.\end{aligned}$$

We now obtain from (11.18) that

$$\|h_n\|_X \leq \|x_n^\delta - x^\dagger\|_X \leq \|x_0 - x^\dagger\|_X,$$

which together with the discrepancy principle yields for $n = N - 2$ that

$$\begin{aligned}\tau\delta \leq \|F(x_{N-1}^\delta) - y^\delta\|_Y &\leq \frac{\sigma + c\|x_0 - x^\dagger\|_X}{1 - c\|x_0 - x^\dagger\|_X}\|F(x_{N-2}^\delta) - y^\delta\|_Y \\ &\leq \left(\frac{\sigma + c\|x_0 - x^\dagger\|_X}{1 - c\|x_0 - x^\dagger\|_X}\right)^{N-1}\|F(x_0) - y^\delta\|_Y.\end{aligned}$$

For $\|x_0 - x^\dagger\|_X$ sufficiently small, the term in parentheses is strictly less than 1, and taking the logarithm shows the desired bound on N. □

If the noise level δ is small, $O(1 + |\log\delta|)$ is a significantly smaller bound than $O(\delta^{-2})$ (for comparable constants, which however cannot be assumed in general), and therefore the Levenberg–Marquardt method can be expected to terminate much earlier than the Landweber iteration. On the other hand, each step is more involved since it requires the solution of a linear system. Which of the two methods is faster in practice (as measured by actual time) depends on the individual inverse problem.

We now consider (local) convergence for noisy data.

Theorem 11.12. *Let Assumption 11.10 hold. If $\|x_0 - x^\dagger\|_X$ is sufficiently small, then $x_n \to \bar{x}$ with $F(\bar{x}) = y$ as $n \to \infty$.*

Proof. From (11.23) for $x = x_0$ and $\tilde{x} = x^\dagger$, we directly obtain that

$$\|F(x_0) - y - F'(x_0)(x_0 - x^\dagger)\|_Y \leq c\|x_0 - x^\dagger\|_X\|F(x_0) - y\|_Y.$$

For $\|x_0 - x^\dagger\|_X$ sufficiently small we then have that $\gamma := \sigma(c\|x_0 - x^\dagger\|_X)^{-1} > 1$ and thus that (11.17) holds. We can thus apply Lemma 11.9 to deduce that $\|x_1 - x^\dagger\|_X \leq \|x_0 - x^\dagger\|_X$. Hence, $x_1 \in B_{2r}(x_0)$ and thus $\|x_1 - x^\dagger\|_X$ is sufficiently small as well. By induction, we then obtain the well-posedness of the iteration and the monotonicity of the error for all $n \in \mathbb{N}$. As in the proof of Theorem 11.11, rearranging and summing yields that

$$\sum_{n=0}^{\infty} \|F(x_n) - y\|_Y^2 \le \frac{\gamma M^2}{2(\gamma-1)(1-\sigma)\sigma}\|x_0 - x^\dagger\|_X < \infty$$

and hence that $F(x_n) \to y$ as $n \to \infty$.

The remainder of the proof proceeds analogously to that of Theorem 11.5. We set $e_n := x_n - x^\dagger$ and consider

$$\|e_m - e_n\|_X \le \|e_m - e_k\|_X + \|e_k - e_n\|_X$$

for any $m \ge n$ and $k \in \{n, \dots, m\}$ chosen according to (11.12). The Féjer monotonicity from Lemma 11.9 again shows that $\|e_n\|_X \to \varepsilon$ for some $\varepsilon \ge 0$ as $n \to \infty$, requiring us to only look at the mixed terms. Using (11.20) and the parameter choice (11.16), we obtain that

$$\begin{aligned}
(e_k - e_m \mid e_k)_X &= \sum_{j=k}^{m-1} -\left(x_{j+1} - x_j \,\middle|\, x_k - x^\dagger\right)_X \\
&= \sum_{j=k}^{m-1} -\alpha_j^{-1}\left(y - F(x_j) - F'(x_j)(x_{j+1} - x_j) \,\middle|\, F'(x_j)(x_k - x^\dagger)\right)_Y \\
&\le \sum_{j=k}^{m-1} \alpha_j^{-1}\|y - F(x_j) - F'(x_j)(x_{j+1} - x_j)\|_Y \|F'(x_j)(x_k - x^\dagger)\|_Y \\
&= \sum_{j=k}^{m-1} \sigma\alpha_j^{-1}\|F(x_j) - y\|_Y \|F'(x_j)(x_k - x^\dagger)\|_Y.
\end{aligned}$$

For the second term, we use (11.23) and set $\eta := c\|x_0 - x^\dagger\|_X \ge c\|x_j - x^\dagger\|_X$ for all $j \ge 0$ to arrive at

$$\begin{aligned}
\|F'(x_j)(x_k - x^\dagger)\|_Y &\le \|F(x_k) - y\|_Y + \|y - F(x_j) - F'(x_j)(x^\dagger - x_j)\|_Y \\
&\qquad + \|F(x_j) - F(x_k) - F'(x_j)(x_j - x_k)\|_Y \\
&\le \|F(x_k) - y\|_Y + c\|x_j - x^\dagger\|_X \|F(x_j) - y\|_Y \\
&\qquad + c\|x_j - x_k\|_X \|F(x_j) - F(x_k)\|_Y \\
&\le (1 + 5\eta)\|F(x_j) - y\|_Y,
\end{aligned}$$

where we have again used multiple productive zeros as well as (11.12).

We can now apply (11.18) to obtain that

$$\begin{aligned}
(e_k - e_m \mid e_k)_X &\le \sum_{j=k}^{m-1} (1 + 5\eta)\sigma\alpha_j^{-1}\|F(x_j) - y\|_Y^2 \\
&\le \sum_{j=k}^{m-1} \frac{\gamma(1 + 5\eta)}{2\sigma(\gamma - 1)}\left(\|e_j\|_X^2 - \|e_{j+1}\|_X^2\right) \\
&= \frac{\gamma(1 + 5\eta)}{2\sigma(\gamma - 1)}\left(\|e_k\|_X^2 - \|e_m\|_X^2\right) \to 0
\end{aligned}$$

as $n \to \infty$ due to the convergence of $\|e_n\|_X \to \varepsilon$. We similarly deduce that

$$(e_k - e_n \mid e_k)_X \leq \frac{\gamma(1+5\eta)}{2\sigma(\gamma-1)} \left(\|e_n\|_X^2 - \|e_k\|_X^2\right) \to 0$$

as $n \to \infty$, which again implies that $\{e_n\}_{n\in\mathbb{N}}$ and hence that $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence. The claim now follows since $F(x_n) \to y$. □

We now have almost everything at hand to apply Lemma 11.1 and show the convergence of the Levenberg–Marquardt method for noisy data $y^\dagger \in Y$.

Theorem 11.13. *Let Assumption 11.10 hold. If $\|x_0 - x^\dagger\|_X$ is sufficiently small, then $x^\delta_{N(\delta,y^\delta)} \to \bar{x}$ with $F(\bar{x}) = y$ as $\delta \to 0$.*

Proof. It remains to verify the continuity condition (11.3b). Since F is assumed to be continuous differentiable, $F'(x^\dagger)^*F'(x^\dagger) + \alpha\,\mathrm{Id}$ is continuous. By the Inverse Function Theorem (e.g., [Renardy & Rogers 2004, Theorem 10.4]), there thus exists a sufficiently small neighborhood of $x^\dagger$ where $(F'(x)^*F'(x) + \alpha\,\mathrm{Id})^{-1}$ is continuous as well. For fixed $n \in \mathbb{N}$, the right-hand side of (11.15) is therefore continuous in x_n, which implies the condition (11.3b) and hence the claimed convergence. □

Under a source condition and with a suitable a priori choice of α_n and $N = N(\delta)$, it is possible to show (logarithmic) convergence rate as $\delta \to 0$; see [Kaltenbacher, Neubauer & Scherzer 2008, Theorem 4.7].

11.3 ITERATIVELY REGULARIZED GAUSS–NEWTON METHOD

We finally consider the following version of the Levenberg–Marquardt method which was proposed in [Bakushinskiĭ 1992]: Set $x_{n+1} = x_n + h_n$ where now h_n is the solution of the minimization problem

$$\min_{h\in X} \frac{1}{2}\|F'(x_n)h + F(x_n) - y\|_Y^2 + \frac{\alpha_n}{2}\|h + x_n - x_0\|_X^2. \tag{11.26}$$

By Lemma 6.3, this is equivalent to the explicit iteration known as the *iteratively regularized Gauß–Newton method*:

$$x_{n+1} = x_n + \left(F'(x_n)^*F'(x_n) + \alpha_n\,\mathrm{Id}\right)^{-1}\left(F'(x_n)^*(y - F(x_n)) + \alpha_n(x_0 - x_n)\right). \tag{11.27}$$

Note that the only difference to the Levenberg–Marquardt method is the additional term on the right-hand side. Similarly, comparing (11.26) to (11.14), the former has $x_n + h_n - x_0 =$

$x_{n+1} - x_0$ in the regularization term. The point is that this allows interpreting x_{n+1} directly as the minimizer of the *linearized* Tikhonov functional

$$\min_{x\in X} \frac{1}{2}\|F'(x_n)(x-x_n) + F(x_n) - y\|_Y^2 + \frac{\alpha_n}{2}\|x - x_0\|_X^2,$$

and hence to use the properties of linear Tikhonov regularization for the analysis. In practice, this method also shows better stability since the explicit regularization of x_{n+1} prevents unchecked growth through the constant addition of (bounded) increments h_n.

As for the Levenberg–Marquardt method, one can now show (under some conditions on the nonlinearity) that this iteration is well-defined and converges for exact as well as noisy data; see [Kaltenbacher, Neubauer & Scherzer 2008, Theorem 4.2]. Instead, we will only show convergence rates for an a priori choice of α_n and $N(\delta)$. To make use of the results for linear Tikhonov regularization from Chapter 6, we assume that F is Fréchet differentiable and completely continuous such that $F'(x)$ is compact for all x by Theorem 9.6. Specifically, we make the following assumptions.

Assumption 11.14. Let $F : U \to Y$ be continuously differentiable and completely continuous, and let $x^\dagger$ be an x_0-minimum norm solution. Assume further that

(i) F' is Lipschitz continuous with constant L;

(ii) there exists a $w \in X$ with $x^\dagger - x_0 = |F'(x^\dagger)|^\nu w$ and $\|w\|_X \leq \rho$ for some $\nu \in [1,2]$ and $\rho > 0$;

We first show that the regularization error satisfies a quadratic recursion.

Lemma 11.15. *Let Assumption 11.14 hold. If the stopping index $N(\delta)$ and α_n, $1 \leq n \leq nN(\delta)$, are chosen such that*

$$\alpha_{N(\delta)}^{(\nu+1)/2} \leq \tau\delta \leq \alpha_n^{(\nu+1)/2} \qquad \text{for all } n < N(\delta) \tag{11.28}$$

and some $\tau > 0$,

$$\|x_{n+1}^\delta - x^\dagger\|_X \leq \left(C_\nu\rho + \tau^{-1}\right)\alpha_n^{\nu/2} + L\rho\left(C_\nu\alpha_n^{(\nu-1)/2} + \|F'(x^\dagger)\|_{\mathbb{L}(X,Y)}^{\nu-1}\right)\|x_n^\delta - x^\dagger\|_X + \frac{L}{2\alpha_n^{1/2}}\|x_n^\delta - x^\dagger\|_X^2 \qquad \text{for all } n < N(\delta).$$

Proof. Using the iteration and rearranging appropriately, we split the regularization error $x_{n+1} - x^\dagger$ into three components that we then estimate separately. We set $K_n := F'(x_n^\delta)$ as well as $K := F'(x^\dagger)$ and write

$$
\begin{aligned}
x_{n+1}^\delta - x^\dagger &= x_n^\delta - x^\dagger + \left(K_n^*K_n + \alpha_n \operatorname{Id}\right)^{-1} \left(K_n^*(y^\delta - F(x_n^\delta)) + \alpha_n(x_0 - x_n^\delta)\right)\\
&= \left(K_n^*K_n + \alpha_n \operatorname{Id}\right)^{-1} \left(\alpha_n(x_0 - x^\dagger) + K_n^* \left(y^\delta - F(x_n^\delta) + K_n(x_n^\delta - x^\dagger)\right)\right)\\
&= \left[\alpha_n \left(K^*K + \alpha_n \operatorname{Id}\right)^{-1} (x_0 - x^\dagger)\right] + \left[\left(K_n^*K_n + \alpha_n \operatorname{Id}\right)^{-1} K_n^*(y^\delta - y)\right]\\
&\quad + \Big[\left(K_n^*K_n + \alpha_n \operatorname{Id}\right)^{-1} K_n^* \left(F(x^\dagger) - F(x_n^\delta) + K_n(x_n^\delta - x^\dagger)\right)\\
&\qquad + \alpha_n \left(K_n^*K_n + \alpha_n \operatorname{Id}\right)^{-1} (K_n^*K_n - K^*K)\left(K^*K + \alpha_n \operatorname{Id}\right)^{-1} (x_0 - x^\dagger)\Big]\\
&=: [e_1] + [e_2] + [e_{3a} + e_{3b}].
\end{aligned}
$$

We first estimate the "approximation error" e_1. Since K is compact, we obtain from Lemma 6.3 the representation $(K^*K + \alpha \operatorname{Id})^{-1}x = \varphi_\alpha(K^*K)x$ for $\varphi_\alpha(\lambda) = (\lambda + \alpha)^{-1}$. Together with the source condition, this implies for all $\nu \leq \nu_0 = 2$ that

$$
\begin{aligned}
\|e_1\|_X &= \|\alpha_n \left(K^*K + \alpha_n \operatorname{Id}\right)^{-1} (x_0 - x^\dagger)\|_X\\
&= \|\alpha_n \varphi_{\alpha_n}(K^*K)(K^*K)^{\nu/2} w\|_X\\
&\leq \sup_{\lambda \in (0,\kappa]} \frac{\alpha_n \lambda^{\nu/2}}{\lambda + \alpha_n}\|w\|_X = \sup_{\lambda \in (0,\kappa]} \omega_\nu(\alpha_n)\|w\|_X\\
&\leq C_\nu \alpha_n^{\nu/2}\rho
\end{aligned}
$$

as shown in Chapter 6.

For the "data error" e_2, we also use the estimates from Chapter 6 together with the a priori choice of α_n to obtain for all $n < N(\delta)$ that

$$
\begin{aligned}
\|e_2\|_X &= \| \left(K_n^*K_n + \alpha_n \operatorname{Id}\right)^{-1} K_n^*(y^\delta - y)\|_X\\
&\leq \|\varphi_{\alpha_n}(K_n^*K_n)K_n^*\|_{\mathbb{L}(Y,X)}\|y^\delta - y\|_Y\\
&\leq \frac{1}{\sqrt{\alpha_n}}\delta \leq \tau^{-1}\alpha_n^{\nu/2}.
\end{aligned}
$$

The "nonlinearity error" $e_{3a} + e_{3b}$ is again estimated separately. For the first term, we use the Lipschitz condition and Lemma 9.5 to bound

$$
\begin{aligned}
\|e_{3a}\|_X &:= \| \left(K_n^*K_n + \alpha_n \operatorname{Id}\right)^{-1} K_n^* \left(F(x^\dagger) - F(x_n^\delta) + K_n(x_n^\delta - x^\dagger)\right) \|_X\\
&\leq \|\varphi_{\alpha_n}(K_n^*K_n)K_n^*\|_{\mathbb{L}(Y,X)}\|F(x^\dagger) - F(x_n^\delta) - F'(x_n^\delta)(x^\dagger - x_n^\delta)\|_Y\\
&\leq \frac{1}{\sqrt{\alpha_n}}\frac{L}{2}\|x_n^\delta - x^\dagger\|_X^2.
\end{aligned}
$$

For the second term, we use the identity

$$
K_n^*K_n - K^*K = K_n^*(K_n - K) + (K_n^* - K^*)K
$$

as well as the Lipschitz continuity of $F'(x)$ and the source condition to estimate similarly as above

$$\begin{aligned}
\|e_{3b}\|_X &:= \|\alpha_n \left(K_n^*K_n + \alpha_n \operatorname{Id}\right)^{-1} \left(K_n^*K_n - K^*K\right) \left(K^*K + \alpha_n \operatorname{Id}\right)^{-1} (x_0 - x^\dagger)\|_X \\
&\leq \|\varphi_{\alpha_n}(K_n^*K_n)K_n^*\|_{\mathbb{L}(Y,X)} \|K - K_n\|_{\mathbb{L}(X,Y)} \|\alpha_n \varphi_{\alpha_n}(K^*K)(K^*K)^{\nu/2} w\|_X \\
&\quad + \|\alpha_n \varphi_{\alpha_n}(K_n^*K_n)\|_{\mathbb{L}(X,X)} \|K_n - K\|_{\mathbb{L}(X,Y)} \|K\varphi_{\alpha_n}(K^*K)(K^*K)^{1/2}\|_{\mathbb{L}(X,Y)} \\
&\qquad \cdot \|(K^*K)^{(\nu-1)/2} w\|_X \\
&\leq \frac{1}{\sqrt{\alpha_n}} L\|x^\dagger - x_n^\delta\|_X \, C_\nu \alpha_n^{\nu/2} \rho + \sup_{\lambda \in (0,\kappa]} \frac{\alpha_n}{\alpha_n + \lambda} L\|x_n^\delta - x^\dagger\| \, \|K\|_{\mathbb{L}(X,Y)}^{\nu-1} \rho \\
&\leq L\rho \left(C_\nu \alpha_n^{(\nu-1)/2} + \|K\|_{\mathbb{L}(X,Y)}^{\nu-1}\right) \|x_n^\delta - x^\dagger\|_X,
\end{aligned}$$

where we have used $\|K^*\|_{\mathbb{L}(Y,X)} = \|K\|_{\mathbb{L}(X,Y)}$ and – applying Lemma 3.13 (iii) – the inequality

$$\|K\varphi_\alpha(K^*K)(K^*K)^{1/2}\|_{\mathbb{L}(X,Y)} = \|(K^*K)^{1/2}\varphi_\alpha(K^*K)(K^*K)^{1/2}\|_{\mathbb{L}(X,X)} \leq \sup_{\lambda \in (0,\kappa]} \frac{\lambda}{\lambda + \alpha} \leq 1.$$

Combining the separate estimates yields the claim. □

If the initial error is small enough, we obtain from this the desired error estimate.

Theorem 11.16. *Let Assumption 11.14 hold for $\rho > 0$ sufficiently small and $\tau > 0$ sufficiently large. Assume further that $\alpha_0 \leq 1$ and*

$$1 < \frac{\alpha_n}{\alpha_{n+1}} \leq q \qquad \text{for some } q > 1.$$

Then we have for exact data (i.e., $\delta = 0$) that

$$\|x_n - x^\dagger\|_X \leq c_1 \alpha_n^{\nu/2} \qquad \text{for all } n \in \mathbb{N} \tag{11.29}$$

and for noisy data that

$$\|x_{N(\delta)}^\delta - x^\dagger\|_X \leq c_2 \delta^{\frac{\nu}{\nu+1}} \qquad \text{as } \delta \to 0. \tag{11.30}$$

Proof. Lemma 11.15 shows that $\xi_n := \alpha_n^{-\nu/2} \|x_n^\delta - x^\dagger\|_X$ satisfies the quadratic recursion

$$\xi_{n+1} \leq a + b\xi_n + c\xi_n^2$$

with

$$a := q^{\nu/2}(C_\nu \rho + \tau^{-1}), \qquad b := q^{\nu/2} L\rho \left(C_\nu + \|F'(x^\dagger)\|_{\mathbb{L}(X,Y)}^{\nu-1}\right), \qquad c := q^{\nu/2}\frac{L}{2}\rho,$$

where we have used that $\nu \geq 1$ and hence that $\alpha_n^{-1/2} \leq \alpha_n^{-\nu/2}$ and $\alpha_n^{\nu/2} < \alpha_0^{\nu/2} \leq 1$. Clearly we can make a, b and c arbitrarily small by choosing ρ sufficiently small and τ sufficiently large. Let now t_1, t_2 be the solutions of the fixed-point equation $a + bt + ct^2 = t$, i.e.,

$$t_1 = \frac{2a}{1 - b + \sqrt{(1-b)^2 - 4ac}}, \qquad t_2 = \frac{1 - b + \sqrt{(1-b)^2 - 4ac}}{2c}.$$

For c sufficiently small, t_2 can be made arbitrarily large; in particular, we can assume that $t_2 \geq \xi_0$. Furthermore, the source condition yields $\|x_0 - x^\dagger\|_X \leq \|F'(x^\dagger)\|_{\mathbb{L}(X,Y)}^\nu \rho$, and hence we can guarantee that $x_0 \in B_r(x^\dagger) \subset U$ for some $r > 0$ by choosing ρ sufficiently small.

We now show by induction that

$$\xi_n \leq \max\{t_1, \xi_0\} =: C_\xi \qquad \text{for all } n \leq N(\delta). \tag{11.31}$$

For $n = 0$, this claim follows straight from the definition; we thus assume that (11.31) holds for some fixed $n < N(\delta)$. Then we have in particular that $\xi_n \leq \xi_0$, and the definition of ξ_n together with the assumptions that $\alpha_n \leq \alpha_0 \leq 1$ and $\nu \geq 1$ imply that

$$\|x_n^\delta - x^\dagger\|_X \leq \alpha_n^{\nu/2} \alpha_0^{-\nu/2} \|x_0 - x^\dagger\|_X \leq r$$

and hence that $x_n^\delta \in B_r(x^\dagger) \subset U$. This shows that the iteration (11.27) is well-defined and that we can apply Lemma 11.15. We now distinguish two cases in (11.31):

(i) $\xi_n \leq t_1$: Then we have by $a, b, c \geq 0$ and the definition of t_1 that

$$\xi_{n+1} \leq a + b\xi_n + c\xi_n^2 \leq a + bt_1 + bt_1^2 = t_1.$$

(ii) $t_1 < \xi_n \leq \xi_0$: Since we have assumed that $t_2 \geq \xi_0$, it follows that $\xi_n \in (t_1, t_2]$, and $a + (b-1)t + ct^2 \leq 0$ for $t \in [t_1, t_2]$ due to $c \geq 0$ implies that

$$\xi_{n+1} \leq a + b\xi_n + c\xi_n^2 \leq \xi_n \leq \xi_0.$$

In both cases, we have obtained (11.31) for $n + 1$.

For $\delta = 0$ we have $N(0) = \infty$, and (11.31) implies that

$$\|x_n - x^\dagger\|_X \leq \alpha_n^{\nu/2} C_\xi \qquad \text{for all } n \in \mathbb{N},$$

yielding (11.29) with $c_1 := C_\xi$. For $\delta > 0$, (11.31) for $n = N(\delta)$ together with the parameter choice (11.28) implies that

$$\|x_{N(\delta)} - x^\dagger\|_X \leq \alpha_{N(\delta)}^{\nu/2} C_\xi \leq (\tau\delta)^{\frac{\nu}{\nu+1}} C_\xi,$$

yielding (11.30) with $c_2 := C_\xi \tau^{\frac{\nu}{\nu+1}}$. □

In a similar way (albeit with a bit more effort), it is also possible to derive convergence rates (up to the saturation $\nu_0 - 1 = 1$) if the stopping index is chosen according to the discrepancy principle, see [Kaltenbacher, Neubauer & Scherzer 2008, Theorem 4.13].

BIBLIOGRAPHY

H. W. Alt (2016), *Linear Functional Analysis, An Application-Oriented Introduction*, Universitext, Springer, London, doi: 10.1007/978-1-4471-7280-2.

R. Andreev, P. Elbau, M. V. de Hoop, L. Qiu & O. Scherzer (2015), Generalized convergence rates results for linear inverse problems in Hilbert spaces, *Numer. Funct. Anal. Optim.* 36(5), 549–566, doi: 10.1080/01630563.2015.1021422.

A. B. Bakushinskiĭ (1985), Remarks on choosing a regularization parameter using the quasioptimality and ratio criterion, *USSR Comput. Math. Math. Phys.* 24(4), 181–182, doi: 10.1016/0041-5553(84)90253-2.

A. B. Bakushinskiĭ (1992), On a convergence problem of the iterative-regularized Gauss-Newton method, *USSR Comput. Math. Math. Phys.* 32(9), 1503–1509.

F. Bauer & M. A. Lukas (2011), Comparing parameter choice methods for regularization of ill-posed problems, *Mathematics and Computers in Simulation* 81(9), 1795–1841, doi: 10.1016/j.matcom.2011.01.016.

H. Brezis (2010), *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Springer, New York, doi: 10.1007/978-0-387-70914-7.

M. Burger (2007), Inverse Problems, Vorlesungsskript, Institut für Numerische und Angewandte Mathematik, Universität Münster, url: http://wwwmath.uni-muenster.de/num/Vorlesungen/IP_WS07/skript.pdf.

N. Du (2008), Finite-dimensional approximation settings for infinite-dimensional Moore–Penrose inverses, *SIAM J. Numer. Anal.* 46(3), 1454–1482, doi: 10.1137/060661120.

H. W. Engl, M. Hanke & A. Neubauer (1996), *Regularization of Inverse Problems*, vol. 375, Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, doi: 10.1007/978-94-009-1740-8.

H. W. Engl, K. Kunisch & A. Neubauer (1989), Convergence rates for Tikhonov regularisation of nonlinear ill-posed problems, *Inverse Problems* 5(4), 523–540, doi: 10.1088/0266-5611/5/4/007.

M. Hanke, A. Neubauer & O. Scherzer (1995), A convergence analysis of the Landweber iteration for nonlinear ill-posed problems, *Numerische Mathematik* 72(1), 21–37, doi: 10.1007/s002110050158.

E. Hewitt & K. Stromberg (1975), *Real and Abstract Analysis*, Springer-Verlag, New York & Heidelberg, doi: 10.1007/978-3-662-29794-0.

B. Hofmann, B. Kaltenbacher, C. Pöschl & O. Scherzer (2007), A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators, *Inverse Problems* 23(3), 987–1010, doi: 10.1088/0266-5611/23/3/009.

B. Hofmann (1999), *Mathematik inverser Probleme*, Mathematik für Ingenieure und Naturwissenschaftler, B. G. Teubner Verlagsgesellschaft mbH, Stuttgart.

T. Hohage (2000), Regularization of exponentially ill-posed problems, *Numer. Funct. Anal. Optim.* 21(3-4), 439–464, doi: 10.1080/01630560008816965.

T. Hohage (2002), Inverse Problems, Lecture notes, Institut für Numerische und Angewandte Mathematik, Universität Göttingen.

K. Ito & B. Jin (2014), *Inverse Problems: Tikhonov Theory and Algorithms*, vol. 22, Series on Applied Mathematics, World Scientific, Singapore, doi: 10.1142/9789814596206_0001.

B. Kaltenbacher, A. Neubauer & O. Scherzer (2008), *Iterative Regularization Methods for Nonlinear Ill-posed Problems*, vol. 6, Radon Series on Computational and Applied Mathematics, De Gruyter, Berlin, doi: 10.1515/9783110208276.

S. Kindermann (2011), Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems, *Electron. Trans. Numer. Anal.* 38, 233–257, url: http://etna.mcs.kent.edu/volumes/2011-2020/vol38/abstract.php?vol=38&pages=233-257.

S. Kindermann (2016), Projection methods for ill-posed problems revisited, *Comput. Methods Appl. Math.* 16(2), 257–276, doi: 10.1515/cmam-2015-0036.

A. Kirsch (2011), *An Introduction to the Mathematical Theory of Inverse Problems*, 2nd ed., Springer, New York, doi: 10.1007/978-1-4419-8474-6.

L. Landweber (1951), An iteration formula for Fredholm integral equations of the first kind, *Amer. J. Math.* 73, 615–624, doi: 10.2307/2372313.

A. K. Louis (1989), *Inverse und schlecht gestellte Probleme*, Teubner Studienbücher Mathematik, B. G. Teubner, Stuttgart, doi: 10.1007/978-3-322-84808-6.

P. Mathé & B. Hofmann (2008), How general are general source conditions?, *Inverse Problems* 24(1), 015009, 5, doi: 10.1088/0266-5611/24/1/015009.

M. Renardy & R. C. Rogers (2004), *An Introduction to Partial Differential Equations*, 2nd ed., vol. 13, Texts in Applied Mathematics, Springer-Verlag, New York, doi: 10.1007/b97427.

A. Rieder (2003), *Keine Probleme mit inversen Problemen*, Eine Einführung in ihre stabile Lösung, Friedr. Vieweg & Sohn, Braunschweig, doi: 10.1007/978-3-322-80234-7.

O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier & F. Lenzen (2009), *Variational Methods in Imaging*, vol. 167, Applied Mathematical Sciences, Springer, New York, doi: 10.1007/978-0-387-69277-7.

T. Schuster, B. Kaltenbacher, B. Hofmann & K. S. Kazimierski (2012), *Regularization Methods in Banach Spaces*, vol. 10, Radon Series on Computational and Applied Mathematics, De Gruyter, Berlin, doi: 10.1515/9783110255720.

A. N. Tikhonov (1963a), On the regularization of ill-posed problems, *Dokl. Akad. Nauk SSSR* 153, 49–52.

A. N. Tikhonov (1963b), On the solution of ill-posed problems and the method of regularization, *Dokl. Akad. Nauk SSSR* 151, 501–504.

B. von Harrach (2014), Regularisierung Inverser Probleme, Vorlesungsskript, Fachbereich Mathematik, Universität Stuttgart.